

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ADMINISTRAÇÃO
DEPARTAMENTO DE CIÊNCIAS ADMINISTRATIVAS**

DANIEL FREDO RODRIGUES

**UM ESTUDO SOBRE O GERENCIAMENTO DE CHURN E A FIDELIZAÇÃO DE
CLIENTES EM UMA EMPRESA DE TELECOMUNICAÇÃO**

**Porto Alegre
2020**

DANIEL FREDO RODRIGUES

**UM ESTUDO SOBRE O GERENCIAMENTO DE CHURN E A FIDELIZAÇÃO DE
CLIENTES EM UMA EMPRESA DE TELECOMUNICAÇÃO**

Trabalho de conclusão de curso de graduação apresentado ao Departamento de Ciências Administrativas da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de Bacharel em Administração.

Orientador: Prof. Dr. Leonardo Nicolao

Porto Alegre

2020

DANIEL FREDO RODRIGUES

**UM ESTUDO SOBRE O GERENCIAMENTO DE CHURN E A FIDELIZAÇÃO DE
CLIENTES EM UMA EMPRESA DE TELECOMUNICAÇÃO**

Trabalho de conclusão de curso de graduação apresentado ao Departamento de Ciências Administrativas da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de Bacharel em Administração.

CONCEITO:

APROVADO EM:

Banca examinadora:

Prof. Dr. Leonardo Nicolao
(Orientador)
(UFRGS)

Prof. Dr. Hugo Fridolino Müller Netto
(UFRGS)

AGRADECIMENTOS

Gratidão! Esta é a palavra que melhor define os meus sentimentos neste momento. Gratidão a Deus, a vida, a minha família, aos amigos e todos que direta ou indiretamente contribuíram para minha jornada acadêmica.

Ao meu orientador professor Dr. Leonardo Nicolao, pela constante motivação, orientação e instrução desde a discussão do projeto de pesquisa até a finalização desse Trabalho de Conclusão de Curso.

Um agradecimento especial à professora Dr^a Mariana Baldi, a primeira professora na graduação que com paciência me orientou, ensinou e me deu forças principalmente no período inicial da faculdade.

Agradeço às minhas Filhas Rafaela e Julia que nesses últimos anos compartilharam comigo o privilégio que é ser estudante da UFRGS. O apoio e companheirismo em todas as circunstâncias foi fundamental para que este ciclo fosse finalizado. Agradeço à minha esposa Rita, que sempre esteve ao meu lado e dividiu comigo todos os momentos até chegarmos aqui.

Por fim, agradeço à Empresa Tri Telecom pelo apoio no tema e pela disponibilização dos meios.

RESUMO

A perda ou abandono de clientes (*churn*), representa impactos significativos nos resultados das organizações. O âmbito desse estudo é uma empresa gaúcha do ramo de telecomunicações, sendo esse um dos segmentos de mercado onde os fornecedores de serviços orientam, cada vez mais, as suas decisões estratégicas na gestão do cliente e na sua satisfação, com a finalidade de promover e consolidar a fidelização deste com a organização. O objetivo desse estudo é, recorrendo à utilização de informações constante em bases de dados da empresa Tri Telecom, aprofundar o conhecimento para identificação de clientes que tendem ao fenômeno do *churn* e criar um modelo que possibilite a antecipação da ocorrência do cancelamento, permitindo a ação por parte da empresa por meio de campanhas de retenção. O tratamento estatístico dos dados e o desenvolvimento do modelo preditivo de *churn* foram realizados por meio do *software* IBM SPSS Statistics. Dentre as técnicas de análise preditiva que podem ser utilizadas na execução dos modelos preditivos utilizou-se a Regressão Logística, tendo em vista a sugestão verificada na literatura de que a metodologia é promissora na previsão de *customer churn*. O trabalho evidencia que o desenvolvimento de ações de retenção pode contribuir significativamente nos resultados da empresa.

Palavras chave: Retenção de clientes, Modelagens preditivas, Regressão logística, SPSS Statistics.

ABSTRACT

The loss or abandonment of customers (churn) represents significant impacts on the results of organizations. The context of this study is a Rio Grande do Sul telecommunications company, which is one of the market segments where service providers increasingly guide their strategic decisions in customer management and satisfaction, with the aim of promoting and consolidate its loyalty with the organization. The objective of this study is, using the information contained in Tri Telecom's databases, to deepen the knowledge to identify customers who tend to the churn phenomenon and to create a model that allows the anticipation of the occurrence of the cancellation, allowing the action by the company through retention campaigns. The statistical treatment of data and the development of predictive churn models were performed using the IBM SPSS Statistics software. Among the data mining techniques that can be used in the execution of predictive models, Logistic Regression was used, in view of the suggestion found in the literature that the methodology is promising in the forecast of customer churn. . The work shows that the development of retention actions can significantly contribute to the company's results.

Keywords: *Customer retention, Predictive modeling, Logistic regression, SPSS Statistics.*

LISTA DE FIGURAS

Figura 1 – Caracterização de <i>Churn</i> em telecom.....	26
Figura 2 – Representação de uma função logística.....	37
Figura 3 – Exemplo de uma curva ROC.....	38
Figura 4 – Exemplo de uma matriz de confusão.....	39
Figura 5 – Curva ROC.....	42

LISTA DE TABELAS E QUADROS

Quadro 1 – A Gestão do relacionamento com o cliente em detalhes.....	22
Quadro 2 – Descrição das variáveis independentes.....	33
Quadro 3 – Métodos do modelos preditivos.....	34
Quadro 4 – Principais Características da regressão logística.....	36
Tabela 1 – Estatística de colinearidade.....	41
Tabela 2 – Área sob a curva.....	41
Tabela 3 – Matriz de confusão.....	42
Tabela 4 – Resumo do processamento do caso.....	43
Tabela 5 – Bloco 0: Bloco inicial.....	45
Tabela 6 – B Bloco 1: Método = Enter.....	45
Tabela 7 _ Tabela de contingência Hosmer e Lemeshow.....	47
Tabela 8 – Tabela de contingência Hosmer	48
Tabela 9 – Tabela 9 Tipo_plano X <i>Churn</i>	49
Tabela 10 – Tabela 10 Incidência_cobrança X <i>Churn</i>	49

LISTA DE GRÁFICOS

Gráfico 1 – Setores da Economia no valor agregado.....	14
Gráfico 2 – Quantidade mensal de rescisões.....	18
Gráfico 3 – Perda mensal de arrecadação.....	18
Gráfico 4 – Clientes cancelados 2020.....	43

LISTA DE SIGLAS

BVP	Busines Value Parteners
CRM	Customer Relationship Management
CSV	Comma Separated Values
GM	Gestão do Conhecimento
IBGE	Instituto Brasileiro de Geografia e Estatística
KDD	Extração do Conhecimento em Banco de Dados
OLAP	On Line Analytical Processing
PIB	Produto Interno Bruto

SUMÁRIO

1 INTRODUÇÃO	12
1.1 DELIMITAÇÃO DO TEMA.....	13
1.2 FORMULAÇÃO DO PROBLEMA.....	14
1.3 JUSTIFICATIVA.....	17
1.4 OBJETIVOS.....	19
1.4.1 Objetivo Geral.....	19
1.4.2 Objetivos Específicos.....	19
2 REFERENCIAL TEÓRICO	20
2.1 MARKETING DE RELACIONAMENTO.....	20
2.2 RETENÇÃO DE CLIENTES.....	23
2.3 <i>CHURN</i>	24
2.4 GERENCIAMENTO DE <i>CHURN</i>	27
2.5 ESTUDOS SOBRE GERENCIAMENTO DE <i>CHURN</i>	27
3 METÓDO	30
3.1 CARACTERIZAÇÃO DA PESQUISA.....	30
3.2 DESCRIÇÃO DA BASE DE DADOS.....	30
3.3 DEFINIÇÃO DAS VARIÁVEIS INDEPENDENTES.....	36
3.3.1 Modelos Preditivos.....	36
3.3.2 Métodos de modelos preditivos.....	37
3.3.3 Regressão Logística.....	37
3.4 AVALIAÇÃO DO MODELO PREDITIVO.....	37
3.4.1 Regressão Logística.....	38
3.4.2 Matriz de confusão.....	38
4 ANÁLISE RESULTADOS	39
4.1 VERIFICAÇÃO DE MULTICOLARIIDADE.....	39
4.2 ÍNDICE DE EXATIDÃO DO TESTE.....	39

SUMÁRIO

4.3	ÍNDICE DE AJUSTE DO MODELO (MATRIZ DE CONFUSÃO).....	41
4.4	INTERPRETAÇÃO DOS RESULTADOS DA R.L.....	43
4.4.1	Bloco 1.....	44
4.4.2	Teste Bloco 1.....	45
5	CONSIDERAÇÕES FINAIS.....	47
	REFERÊNCIAS.....	53

1 INTRODUÇÃO

De acordo com Hennig-Thurau et al. (2004), o relacionamento duradouro com clientes é fundamental para a sobrevivência econômica de uma empresa de serviços, dado que o custo para adquirir novos consumidores costuma ser mais alto do que manter os clientes já conquistados, considerando-se também que a concorrência tem aumentado, em especial na área de telecomunicações. Para Pfeifer (2005), mesmo não havendo concordância com relação à razão dos custos para atração e retenção de clientes, há um consenso quanto ao investimento de esforços e recursos na retenção de clientes. Utilizando modelos preditivos é possível identificar antecipadamente quais clientes podem abandonar a empresa. Segundo Kuhn e Johnson (2013), modelagem preditiva é o processo de criação de um modelo matemático para tentar prever um resultado, podendo esta técnica ser aplicada em diversas áreas.

O *churn*, também conhecido como *churn rate* é um indicador que mede a taxa de cancelamento em um determinado período. A predição de *churn*, é um assunto bastante explorado ultimamente. Todavia, em virtude das particularidades de cada empresa e da quantidade de métodos existentes, escolher um algoritmo para resolver o problema de *churn* pode ser uma tarefa bastante morosa, dado que as comparações são normalmente realizadas de forma manual através de métricas de avaliação. Segundo Hossin e Sulaiman (2015), as métricas de avaliação são o ponto-chave para a escolha do algoritmo ideal e, de acordo com Kumar (2012), não existe um algoritmo que seja o mais indicado em todas as situações, portanto, entende-se que a escolha do melhor algoritmo é um ponto crítico para o desenvolvimento da solução, visto que uma escolha inadequada das variáveis que compõem o algoritmo poderá acarretar resultados indesejados. A principal motivação para a presente pesquisa transcorre devido ao setor de telecomunicações junto ao setor bancário ser hoje, o que mais apresenta concorrência, similaridade da oferta de serviços, competitividade, busca por fidelização de clientes e interesse em se evitar o *churn* de clientes (BERSON; SMITH; THEARLING, 2000; PORTER, 2000).

Utilizando a base de dados de uma empresa de Telecomunicações que é provedora de serviços de telefonia e Internet (fibra óptica) para pessoas físicas e jurídicas, é proposto um estudo para definir um algoritmo com melhor desempenho, para realizar a predição de *churn*. O estudo está estruturado da seguinte forma: esta introdução apresenta a delimitação do tema, a formulação do problema, a justificativa e os objetivos e a seguir, são apresentados o referencial teórico, o método da pesquisa e a ponderação dos resultados. O estudo é finalizado com as principais conclusões, implicações e sugestões para futuras pesquisas.

1.1 DELIMITAÇÃO DO TEMA

Neste contexto e, acompanhando a atual situação das empresas de telecomunicação, que apresentam dificuldade em manter seu quadro de clientes, dado que um grande número cancelaram seus contratos ou migraram para concorrência, que o presente estudo está inserido.

A TRI TELECOM, empresa do ramo de telecomunicações, foi fundada em 2008 com o nome de Tríade Tecnologia, oferecendo serviços de *Outsourcing* de banda Larga via rádio, na região metropolitana de Porto Alegre. No ano de 2012 a empresa recebeu da Anatel a outorga que habilitou o fornecimento de telefonia fixa em todo território nacional. No ano de 2014, a Tri consolidou a nova estratégia de *brandig*, adotou seu nome atual e nova identidade visual, utilizados até hoje. O nome curto e simples busca identificação imediata com o público gaúcho, principal área de atuação geográfica da operadora. Após a empresa transferir sua sede para cidade de Porto Alegre está ampliando seu *Market Share* e tem efetuado ações de Marketing para captação de clientes nesse novo território, contudo mesmo com o ingresso de clientes novos a empresa não consegue suprir os clientes perdidos durante os meses. No presente a empresa faz, em média, a ativação de 25 clientes novos por dia, todavia, em média, 5 clientes são perdidos diariamente.

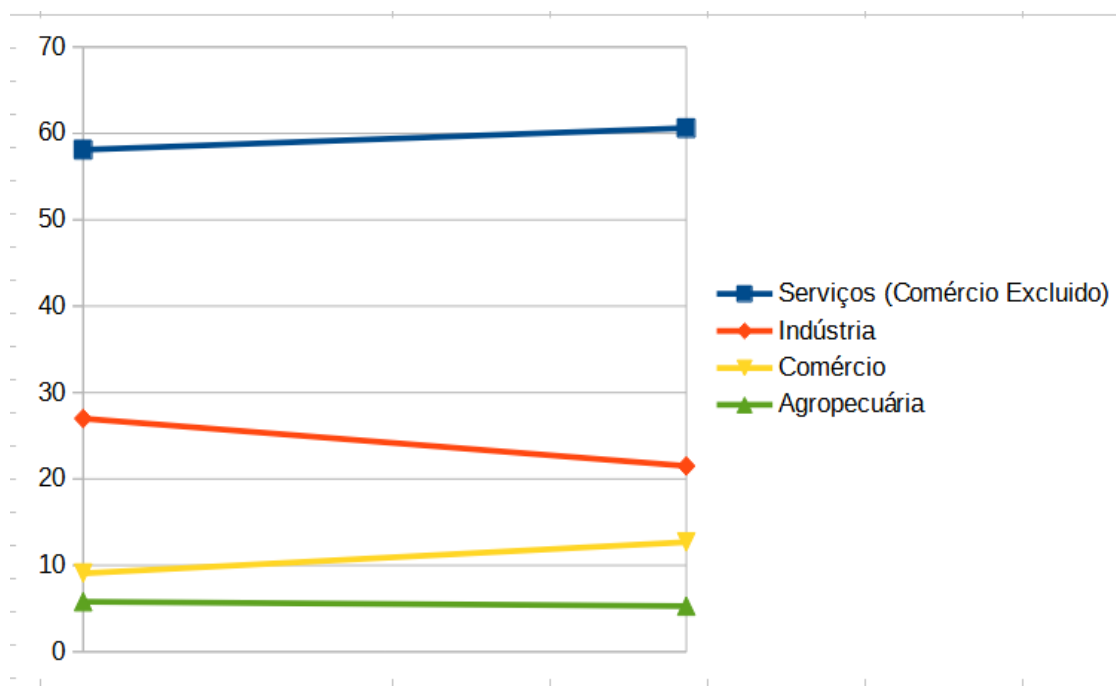
As empresas de telecomunicação, para dimensionar o percentual de clientes que mudaram para concorrência ou cancelaram o contrato utilizam uma metodologia

do marketing de relacionamento denominada gerenciamento de *churn*, a qual fornece informações para que as organizações desenvolvam ações que mitiguem os casos de abandono, bem como propiciam a manutenção de um relacionamento de longo prazo com o cliente.

1.2 FORMULAÇÃO DO PROBLEMA

Considerando o Valor Adicionado das atividades no ano de 2017 o PIB Produto Interno Bruto totalizou R\$ 6.559,9 bilhões, sendo R\$ 5.648,6 bilhões referentes ao Valor Adicionado a preços básicos e R\$ 911,4 bilhões aos Impostos sobre Produtos Líquidos de subsídios, a Agropecuária registrou R\$ 299,5 bilhões, a indústria R\$ 1.212 bilhões o Setor de Comércio e Serviços R\$ 4.137,1 bilhões, o que perfaz uma participação de 73,2% no total de valor adicionado na economia. A importância do setor de serviços fica evidente quando se olha sua parcela na geração da riqueza do país. A participação do setor no valor adicionado passou de 58,1% em 1995 a 60,6% em 2017(Gráfico 1).

Gráfico 1 – Setores da Economia no Valor Agregado



Fonte: IBGE (2017)

Segundo Kotler (1999) serviço é qualquer ato ou desempenho que uma parte pode oferecer a outra e ser essencialmente intangível e não resulta em propriedade, podendo ou não estar ligado a um bem físico quando da sua execução. Conforme Zeithaml e Bitner (2000), serviços são atos, processos e performances, como também todas as atividades econômicas cujo produto não é físico ou construído. Para Troster (1999), os serviços são aquelas atividades que, sem criar objetos materiais, se destinam direta ou indiretamente a satisfazer necessidades humanas. É de suma importância observar que o principal objetivo dos fornecedores de serviços é igual ao de outros setores, a saber: desenvolver e proporcionar ofertas que satisfaçam as necessidades dos consumidores e suas expectativas, garantindo, assim, sua própria sobrevivência econômica segundo afirmam Zeithaml e Bitner (2000).

Atualmente as empresas fornecedoras de serviço tem se deparado com um ambiente competitivo e de constante evolução principalmente na questão tecnológica. Neste contexto os custos para conquistar novos clientes sofreram uma majoração, dado que a concorrência está mais acirrada, e também devido aos avanços tecnológicos, há sempre a necessidade de investimentos em atualização. Kotler (1999), orienta que conquistar novos clientes custa entre 5 a 7 vezes mais que manter os já existentes.

Para Vavra (1993):

O que é consideravelmente mais difícil em tal dinâmica de mercado é manter os clientes comprando regularmente uma marca ou serviço dados os custos decorrentes de conquistar novos clientes. A única maneira de lucrar em tal situação é aumentar o tempo de vida de compra dos clientes atuais. Portanto, a retenção de clientes é, de longe, mais importante do que a atração de clientes (VAVRA, 1993, p. 17)

Diante disso, as principais alternativas para que a empresa obtenha um desempenho de lucratividade são: diminuir os custos de aquisição de clientes e manter a lealdade, mantendo-os satisfeitos e em uma relação duradoura. Para conhecer o valor que o negócio investe para conquistar cada cliente novo o indicador mais utilizado é o de Custo de Aquisição de Cliente, ou *Customer Acquisition Cost*, que, na prática analisa todo o processo de vendas da empresa

comparado com o número de clientes conquistados em determinado período. Para Zeithaml e Bitner(2000), lealdade é uma estratégia cujo objetivo maior é criar valor para os clientes, para funcionários e para a empresa. A lealdade é descrita com a existência de um comprometimento profundo em comprar ou utilizar novamente um serviço no futuro.

No entanto, atualmente conquistar a lealdade dos clientes não é tarefa simples, dado que os clientes estão mais exigentes e têm mais consciência do seu poder de compra e também dos seus direitos, soma-se ainda o fato que há um número maior de empresas que disputam pela preferência de um mesmo cliente. Para suprir essa necessidade foi desenvolvida na disciplina de Marketing, nos anos 90, uma área conhecida como Marketing de Relacionamento, que é um conjunto de ações tomadas pela empresa para criar e manter um relacionamento positivo com os seus clientes. Para reforçar essa relação de proximidade, a empresa oferece benefícios para garantir a fidelidade dos seus clientes.(KOTLER; KELLER, 2012).

De acordo com McKenna (1991), “marketing de relacionamento” ou pós-marketing significa construir e sustentar a infraestrutura dos relacionamentos de clientes. É a integração dos clientes com a empresa, desde o projeto, desenvolvimento de processos industriais e de vendas até o término do relacionamento comercial.

Conforme Bogman (2002, p. 23) “o marketing de relacionamento é essencial ao desenvolvimento da liderança no mercado, à rápida liderança no mercado, à rápida aceitação de novos produtos e serviços e à consecução da fidelidade do consumidor”.

Uma métrica utilizada para descobrir o grau de infidelidade dos clientes, ou seja, mede qual o percentual de clientes que deixaram de consumir os serviços de determinada empresa, para consumir os serviços da concorrência é a *churn rate*, ou somente *churn*. No Brasil esse fenômeno tem afetado principalmente as empresas de telecomunicações. Hoffman e Bateson (1997,p. 372) definiram retenção de *churn* como “processo sistemático de tentar ativamente reter consumidores antes deles desertarem”.

Segundo Cister (2005) são três as categorias de *churn*: involuntário – quando

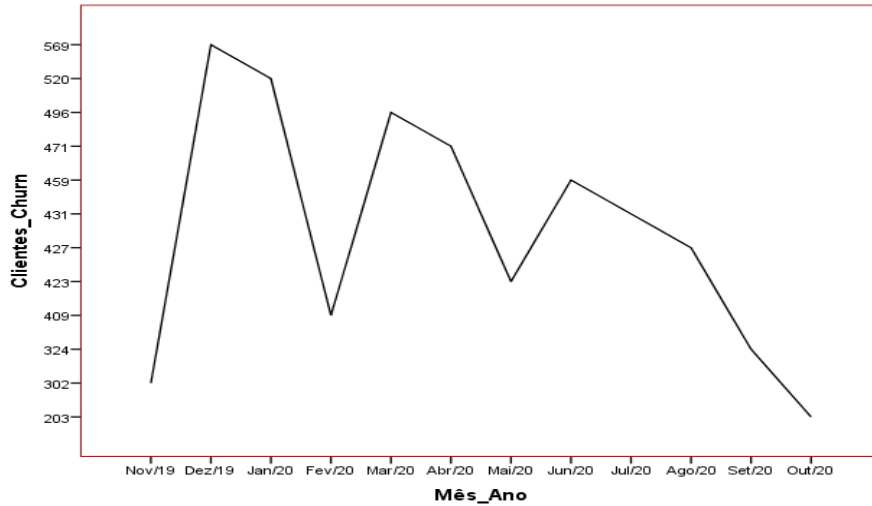
o cliente não cumpre com a sua parte do contrato, deixa de pagar e tem o fornecimento interrompido, voluntário – quando o cliente influenciado por promoções ou campanhas de marketing decide trocar de fornecedor e o inevitável – quando o cliente falece ou muda para uma localidade na qual a empresa não fornece os seus serviços.

De acordo com estudos da BVP - Business Venture Partners, uma taxa aceitável de *churn* seria entre 5 % e 7 % ao ano. No entanto, não há um valor padrão, porque depende da maneira como cada empresa lida com a sua base de clientes. De toda forma, é uma taxa que demonstra como a empresa está lidando com o sucesso do cliente e com a qualidade dos serviços oferecidos. Efetuando uma pesquisa on-line não foi possível identificar qual é taxa média de *churn* apresentada nas empresas de telecomunicação, atualmente, contudo uma pesquisa da TELECO (2007) apresentou uma taxa variando entre 1,8% e 7,2% nas empresas de telefonia celular. Diante disso, o presente estudo coloca como proposição o seguinte questionamento: É possível elaborar um modelo de predição de *churn*, de tal forma que o cliente possa ser retido antes de abandonar a organização?

1.3 JUSTIFICATIVA

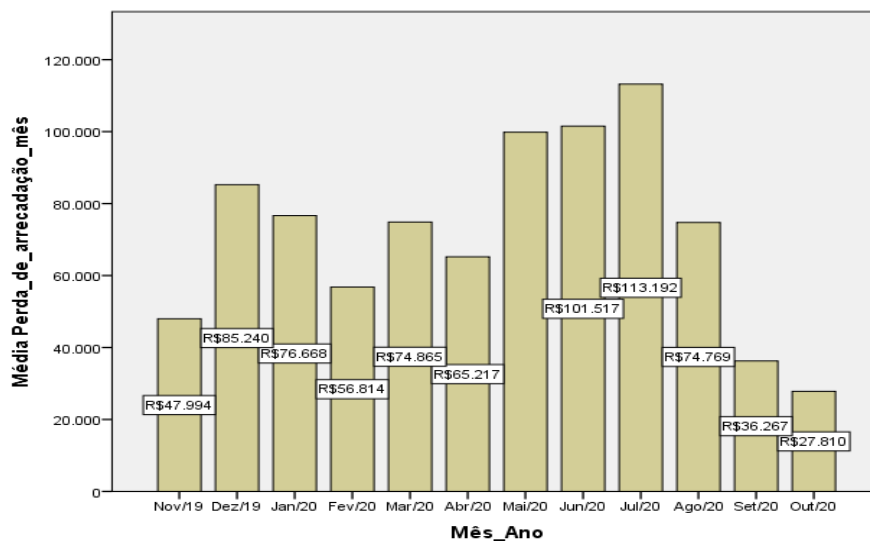
A Tri Telecom está realizando investimento em ações de marketing para captação de novos clientes, contudo para que este sistema tenha eficiência, se faz necessário estabelecer um relacionamento duradouro com os clientes, dado o *payback* do investimento na ativação de um cliente gira em torno de 6 a 7 meses, e somente a partir desse período que a relação proporciona lucro para empresa.

Conforme levantamento preliminar efetuado no banco de dados da Empresa Tri Telecom referente ao período de Nov/2019 a Dez/2020 verificou-se que o número absoluto de clientes *churn* apresentou queda, no entanto, em termos percentuais o índice é considerado alto, dado que a métrica de churn rate em Nov/2020 estava cerca de em 20 %. (Gráfico 2).

Gráfico 2 - Quantidade mensal de rescisões

Fonte: Elaborado pelo autor

Ainda no levantamento preliminar efetuado no banco de dados da Empresa Tri Telecom referente ao período de Nov/2019 a Dez/2020 foi possível verificar que as rescisões dos clientes ocasionaram significativa perda de arrecadação. No gráfico 3 é possível verificar os montantes que não foram arrecadados em virtude dos contratos cancelados.

Gráfico 3 – Perda mensal de Arrecadação

Fonte: Elaborado pelo autor

Esta pesquisa pretende saber quais os motivos que levam os clientes a cancelarem seus contratos de serviço com a Tri Telecom, e com os dados coletados construir um modelo preditivo de *churn*. E se torna relevante dado que a concorrência no mercado de telecomunicações está acirrada e manter os clientes fidelizados é um importante fator de diferenciação.

Além disso, essa pesquisa é uma oportunidade de corroborar com os estudos organizacionais referentes à temática de gerenciamento de *churn* nas empresas de telecomunicações brasileiras, tendo em vista a similaridade entre as organizações.

1.4 OBJETIVOS

Os objetivos para executar o estudo estão classificados em objetivo geral e objetivos específicos.

1.4.1 Objetivo Geral

- Elaborar um modelo de predição de *churn*, para clientes ativos com o propósito de fidelizar os clientes.

1.4.2 Objetivos Específicos

- Adquirir e adequar as bases de clientes;
- Investigar o perfil dos clientes ativos e cancelados em uma empresa de telecomunicação;
- Identificar e analisar as principais razões de *churn* em uma empresa de telecomunicação.

2 REFERENCIAL TEÓRICO

Este capítulo tem pretende abordar alguns conceitos que a literatura apresenta referentes a Marketing de Relacionamento, a Retenção de Clientes, ao *Churn*, Gerenciamento de *Churn*.

2.1 MARKETING DE RELACIONAMENTO

McKenna(1991) descreveu o Marketing de relacionamento como “Novo Marketing”, um marketing baseado em experiências, em que os consumidores definem uma hierarquia de valores, desejos e necessidades, com base em dados empíricos, opiniões e referências obtidas através de propaganda de boca e experiências anteriores com produtos e serviços, as tomadas de decisão de compras são baseadas nessas informações. Para o autor, o Marketing de relacionamento compreende tanto ouvir quanto falar com o cliente, sendo através desse diálogo que as relações são construídas e os produtos são concebidos, adaptados e aceitos. Quando todos os mercados começarem a parecer nichos de mercado com infraestruturas nítidas, o processo de comunicação ficará cada vez mais direto.

O surgimento do Marketing de Relacionamento, bem como o próprio desenvolvimento do Marketing tradicional, teve sua origem pela evolução natural das práticas empresariais e da própria configuração do mercado, que se tornou competitivo(Palmer, 2002; Parvatiyar & Sheth, 2000). Apesar de na sua concepção original a sua abordagem ser voltada para as transações entre organizações, paulatinamente tornou-se proposição para mercados voltados ao consumidor final(O'Malley & Tynan, 2000). Atualmente apresenta-se como a atividade mais importante do Marketing e se preocupa principalmente em como manter a existência e permanência dos consumidores na empresa por mais tempo, mais que em como fazer para obter novos consumidores (REICHHELD et al., 2000).

Conforme Kotler (2003), o CRM é um processo interativo que transforma informações sobre os clientes em relacionamentos positivos com os mesmos. CRM não é apenas um software ou uma atividade exclusiva do setor de atendimento, mas

também é um processo de antecipar e satisfazer as expectativas dos clientes e pode ser definido como um conjunto de estratégias de gerenciamento, através de relacionamento com o cliente, que implica maior lucratividade e ganhos de vantagem competitiva.

Para Kotler e Keller (2006) a gestão do relacionamento com o cliente – CRM, do inglês *Customer Relationship Management* – trata do “gerenciamento cuidadoso de informações detalhadas sobre cada cliente e todos os pontos de contato com ele, de modo a maximizar sua fidelidade”. Complementando, Zeithaml e Bitner (2003) apontam como principal objetivo do CRM “a construção e a manutenção de uma base de clientes comprometidos que sejam rentáveis para a organização”. Assim, gerenciar o relacionamento com o cliente nada mais é do que conhecer suas necessidades, desenvolver novos produtos e oferecer novos serviços a fim de mantê-los satisfeitos e fiéis prolongando e amadurecendo sua permanência na empresa.

Conforme destaca Light (2003), a escolha de uma plataforma tecnológica para operar CRM não é uma tarefa simples, pois, além da escolha do software, mais importante é a identificação das necessidades e do que a empresa espera obter com as práticas relacionais. Todos *stakeholders* devem fazer parte das delimitações das ações relacionais, mas especial as áreas de venda e pós-vendas e os setores de atendimento de clientes. Esta noção pode ser ampliada, tendo em vista além da maximização do contato com os clientes, o cuidado na relação com os parceiros de negócio, fornecedores e no próprio modo de relacionamento da organização com seus funcionários, entendendo que são a interface da organização com seus clientes (GALBREATH, ROGERS, 1999). No entanto, os tipos de CRM são predominantemente classificados nas três seguintes categorias:

CRM operacional: sua principal função é a de automatizar o processo de CRM no intuito de melhorar sua eficiência e produtividade. Pontos de atendimento e suporte aos clientes, como por exemplo, os call centers, automatização da força de vendas e automatização das regras de negócio pertencem a esta categoria;

CRM analítico: trata-se da inteligência do processo de CRM, que consiste na análise e Gestão do Conhecimento (GC) dos clientes. Isso se dá por meio da

modelagem preditiva sobre seu comportamento e padrão de compra, proporcionando uma melhor compreensão e entendimento sobre este comportamento e sobre suas necessidades individuais. Exemplos de ferramentas que fazem parte desta categoria são: *Data Mining*, *Data Warehouse* e o *Online Analytical Processing (OLAP)*;

CRM colaborativo: seu objetivo é gerenciar, integrar e sincronizar os canais de comunicação e os pontos de contato do cliente com a organização. Exemplos de ferramentas do CRM colaborativo são os portais e *websites*, e-mail, redes sociais entre outros. Na tabela a seguir é possível identificar os principais benefícios de implementar uma estratégia de CRM, na prática (Quadro 1).

Quadro 1: A Gestão do relacionamento com o cliente em detalhes

A tecnologia CRM pode ajudar a ...				
Analisar dados sobre receita e custo com o cliente para identificar clientes de alto valor atual e futuro.	Capturar dados relevantes sobre o comportamento de produtos e serviços.	Processar transações mais rapidamente.	Alinhar incentivos e indicadores	Monitorar os índices de perda e retenção de clientes.
	Criar canais de distribuição.	Fornecer informações melhores para linha de frente.	Disseminar sistemas de gestão de conhecimento.	Monitorar os índices de satisfação do atendimento ao cliente.
Direcionar melhor os esforços de marketing direto.	Desenvolver novos modelos de determinação de preços.	Gerenciar com mais eficiência a logística e a cadeia de suprimentos.		
	Construir comunidades	Catalisar o comércio colaborativo.		

Fonte: elaborado pelo autor adaptado de KOTLER, Philip, KELLER, Kevin L. (2012)

O marketing de relacionamento é um processo com atividades contínuas capazes de manter um cliente em situações de pós-vendas, ou seja, a empresa cria vínculos mais sólidos com seus clientes finais direcionando assim para a fidelização e retenção dos mesmos.

2.2 RETENÇÃO DE CLIENTES

No trabalho de Vavra e Pruden (1995) encontramos a afirmação de que a retenção de clientes é a chave para o sucesso no mercado contemporâneo, tornando-se o mais importante componente para a consolidação da participação de mercado sendo direcionada em satisfazer o cliente, em consonância com que encontramos em Rust e Zahorik (1993), contudo, o desafio vai além da preocupação com a simples satisfação do cliente, porque na realidade, é uma nova visão a respeito do que produz lucro para organização e sobre a maneira como se deveria operar nas relações com os clientes e como se deveria negociar (VAVRA; PRUDEN, 1995). Conforme Reinchheld (1993) e Reichheld, Markey Jr. E Hopton (2000) não se pode contestar que os lucros são fundamentais, não com um fim, mas principalmente porque propiciam à empresa a condição de gerar valor bem como criar incentivos para que os clientes permaneçam leais. Como uma etapa decorrente da retenção de cliente, é imperioso lembrar que os benefícios da lealdade geralmente se constituem na razão pela qual um competidor é mais lucrativo que outro (REICHHELD, 1993).

A despeito da lealdade e retenção de clientes sejam apresentadas como certeza de maior participação de mercado, rentabilidade e lucratividade mais elevadas no longo prazo, não há unanimidade quanto ao desenvolvimento desse processo e como se comportam os clientes (McALEXANDER; SCHOUTEN; KOENIG,2002). Nesse sentido, Dowlinge Uncles (1997) corroboram essa ideia, afirmando que, apesar dos benefícios provenientes da retenção e da lealdade de clientes, há evidência de que alguns problemas ocorrem. Conclui-se, portanto, que nem todos os clientes devem ser alvos de esforços concentrados no sentido de retenção e fidelização, pois alguns clientes que mesmo que satisfeitos, podem adotar o comportamento de troca de fornecedores por razões que extrapolam o controle, apesar do empenho da empresa (GANESH;ARNOLD;REYNOLDS,2000).

Hennig-Thurau e Klee (1997), criaram um modelo conceitual chamado Qualidade do Relacionamento (Figura 2), o qual envolve os seguintes construtos como determinantes da retenção de clientes: satisfação de clientes, qualidade

percebida, confiança e comprometimento. Os autores concluíram, com base neste modelo, que um alto nível de qualidade favorece a percepção positiva do cliente em relação aos serviços prestados, resultando em maiores possibilidades do estabelecimento de uma relação duradoura, podendo repercutir positivamente na retenção de clientes. Em conformidade com as ideias dos autores, Verhoef (2003), afirma que a ocorrência de um alto nível de retenção de clientes faz com que a empresa tenha um equilibrado conhecimento da sua carteira de clientes, fomentando melhores relacionamentos entre clientes e prestador de serviços temporalmente.

2.3 CHURN

Strouse (1999) e Ferreira (2005) observam que esse fenômeno do *churn* tem como base o marketing e é uma forma de medir a fidelidade dos clientes, dado que trata da perda de clientes sofrida pela empresa para a concorrência. Observam também que nas relações entre os clientes e as empresas esse fenômeno demonstra exatamente o significado atribuído ao verbo *to churn* da língua inglesa, que pode ser entendido como mexer, rodar, agitar fortemente, bater (FERREIRA, 2005), pois, também há uma agitação de clientes no mercado, trocando de empresas com certa frequência, forçando-as a buscarem novas formas de manter seus clientes no negócio e ainda conquistar os consumidores da concorrência.

Linoff e Berry (2011) consideram o termo *churn* comum na indústria de serviços, com certo destaque para os serviços com alguma semelhança e concorrência – como é o caso do mercado de telecomunicações, seguros, cartões de crédito e serviços financeiros –, em que esse fenômeno tem ampla utilização para medição da taxa de abandono de clientes. Dizer que a taxa de *churn* de uma empresa varia entre 3% e 5% anualmente significa dizer que tal empresa está perdendo de 3% a 5% de seus clientes ativos por ano (FERNANDES NETO; MATTOZO; COSTA, 2008). No entanto, essa taxa significa mais que a perda do cliente, dado que, a empresa precisa repor essa evasão para poder manter sua base

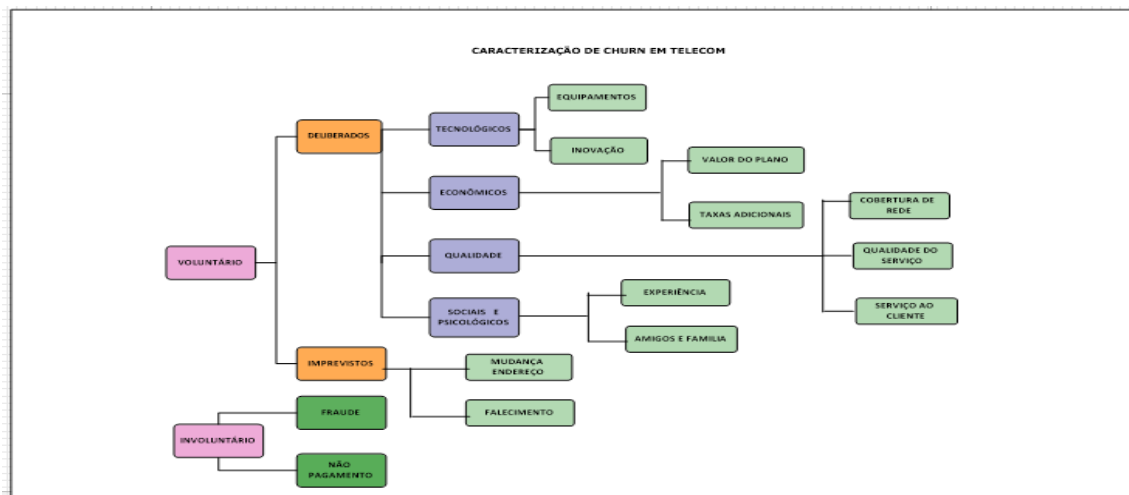
de clientes. Nesse sentido que algumas traduções do termo trazem a definição de rotatividade de clientes, indo além da definição mais simples de perda de clientes (STROUSE, 1999).

Conforme Hadden et al., (2007) há duas categorias de *churn*: involuntário e o voluntário. O involuntário é quando o cliente não cumpre com as suas obrigações financeiras e a empresa deixa de oferecer os produtos ou serviços; e o voluntário é quando o cliente decide não efetuar a compra ou cortar o relacionamento com a empresa, este se divide em acidental, em que o cliente não consegue por algum motivo manter o relacionamento com a empresa (por exemplo, mudança de endereço), e o deliberado que é quando o cliente toma a decisão de trocar de fornecedor, por melhor oferta da concorrência. Segundo Neslin et al, (2006), uma forma de gerenciar o *churn* é prever quais clientes com mais probabilidade de abandonar o relacionamento com a empresa e trabalhar com eles a fim de tentar evitar a ocorrência desse rompimento.

MATTISON (2005) observa que quando se procura compreender os reais motivos que levam um cliente a abandonar o operador, é possível perceber que há pouco esforço no sentido da interpretação do seu perfil, enquanto cliente, tanto na sua motivação inicial para adesão ao serviço, bem como as razões que o levam a querer mudar de operador. Mesmo sendo, o fator preço, a principal motivação na origem do abandono de um cliente, a decisão que o leva a mudar de operador é um processo complexo no qual estão presentes diversos fatores; entre os quais os tecnológicos, econômicos, qualidade, entre outros. Ainda segundo o autor o *churn* pode ser classificado em **Churn Voluntário** e **Churn involuntário**, O *churn* voluntário acontece quando um cliente inicia o processo de encerramento do contrato. O *churn* do tipo voluntário pode ser subdividido em dois grandes grupos: – **Churn Imprevisto**. Esta categoria de *churn* acontece não por ter sido previamente planejado pelo cliente, porque, devido a determinado acontecimento na vida do cliente, este encontra-se obrigado a encerrar o contrato. Este encerramento de contrato é visto como um efeito colateral desse acontecimento.– **Churn Deliberado**. Nesse caso há um processo de tomada de decisão por parte do cliente sustentado em um ou mais fatores, tais como: – Tecnologia - Cliente espera uma nova ou

melhor tecnologia. – Econômicos - Sensibilidade ao preço. – Qualidade - Melhor qualidade dos produtos/serviços. – Sociais - Questões do grupo social onde o cliente está inserido. – Psicológicos - Questões meramente psicológicas (vontade). – **Churn Involuntário**. Neste tipo de *churn* não se verifica uma intenção por parte do cliente em mudar de operador, mas sim um conjunto de circunstâncias que forçam a finalização da relação comercial. Este é o tipo de *churn* mais frequente na maioria das organizações e com maior impacto financeiro. As circunstâncias que geram este *churn* podem ser divididas em duas subcategorias: 1) Fraude – Quando são realizadas atividades proibidas por lei. 2) Churn por não pagamento – Clientes com problemas financeiros. (Figura 1).

Figura1 - Caracterização de Churn em telecom



Fonte: elaborado pelo autor adaptado de Mattison (2005)

Engel, Blackwell e Miniard (2005) chamam a atenção para a necessidade de um tipo de mecanismo de aviso antecipado que possa indicar problemas em tempo que seja suficiente para a tomada de ações corretivas. Os autores alertam que, quando o cliente está na situação de abandono, ou seja, cancelando suas últimas operações com a instituição, já é tarde demais para efetivar medidas de retenção. O motivo pelo qual o *churn* gera impactos financeiros para as empresas ocorre pelo fato do avanço da tecnologia e da globalização aumentarem a concorrência e os custos associados à atração de novos clientes (CASTELLS, 2010).

2.4 GERENCIAMENTO DE *CHURN*

Engel, Blackwell, Miniard (2000, p. 181) estabelecem que para evitar o cancelamento faz-se necessário um “sistema de aviso antecipado” que indique os problemas em tempo hábil para que sejam tomadas ações corretivas, pois quando o cliente tem seu contrato cancelado pela empresa ou solicita o cancelamento, de modo geral é demasiado tarde para que as medidas de retenção surtam efeito. Para isso o *churn* management (Gerenciamento de *churn*) aborda o desenvolvimento de técnicas para mensurar e administrar a taxa de abandono dos clientes para reter consumidores antes mesmo deles abandonarem a empresa e, se isto ocorre, investigar e determinar as razões pelas quais eles a abandonam (IKEDA, 2006, p.397).

Segundo Lejeune (2001) o gerenciamento de *churn* não é somente uma ferramenta de retenção, mas também um desenvolvimento de técnicas que possibilitam às empresas manter seus clientes mais lucrativos e procura elevar a lealdade dos clientes. O autor lembra, também, que existe uma diferença entre reter o cliente e elevar sua lealdade, salientando que o gerenciamento deve manter o foco em clientes e não a toda a base de clientes, por essa razão é de suma importância conhecer cada cliente.

Para um gerenciamento eficaz do *churn* é necessário conhecer a fundo a base de clientes que está na empresa, suas particularidades e desejos. No entanto, é preciso ter um sistema de informação desenvolvido e atualizado para que possa fazer uma análise precisa do banco de dados. Essa análise profunda permite, segundo Cister (2005), não só identificar os clientes que desistiram do serviço, como também identificar os clientes lucrativos e que merecem atenção especial por parte da empresa.

2.5 ESTUDOS SOBRE GERENCIAMENTO DE *CHURN*

Nas últimas décadas, o crescimento da internet, juntamente das tecnologias associadas têm modificado a forma como as empresas e clientes se relacionam.

Comumente a retenção de um cliente custa muitas vezes menos que a aquisição de um novo. Assim sendo estudos de gerenciamento de *churn* e retenção de clientes tem se tornado muito importante para as empresas, e tem se mostrado importante subsídio na gestão. Em um trabalho pioneiro(SCHNEIDER,2016) reuniu 80 estudos onde constatou-se que as técnicas mais utilizadas para construir modelos preditivos de perda de clientes foram a regressão logística, técnicas de *Machine Learning* e árvores de decisão, entre outras, e ainda segundo os resultados obtidos área de telecomunicação foi uma das principais em que os estudos foram aplicados.

Costa et al., (2008). Em seu estudo procuraram analisar os dados de banda largas armazenadas pelas empresas de telefonia, com o objetivo de encontrar a relação entre os dados de qualidade de serviço com a taxa de abandono. Costa et al.,(2008) descreve:

Nesse cenário, foram realizados estudos utilizando-se vários métodos de seleção de variáveis, por meio de regressão linear múltipla, para relacionar os indicadores de qualidade de serviço com a taxa de abandono (Churn). Os resultados obtidos permitem sugerir que os indicadores de qualidade, Tempo Médio de Reparo, Tempo Médio de Instalação, Preventivas e Taxa de Reparo podem influenciar substancialmente no abandono de clientes da planta das operadoras.

Também com o objetivo de analisar o *churn* de uma empresa de telecomunicação, Santos et al., (2012) realizaram um estudo de caso qualitativo, descritivo e exploratório, realizando um levantamento de dados da empresa, e após o estudo realizado e a partir dos resultados, Santos et al., (2012) conclui: “o marketing de relacionamento faz um diferencial na empresa, pois ele ajuda à mesma a crescer e aprimorar seus relacionamentos com seus clientes, pois eles são as riquezas necessárias para o sucesso da organização.”

Com uma abordagem da Mineração de textos baseada na extração de regras de associação que demonstram as percepções e experiências que dos clientes com relação aos serviços prestados por uma empresa de telecomunicações, bem como a classificação dos clientes, Vasques et al., (2017) realizaram um estudo utilizando uma base de dados, contendo ocorrências diversas, registradas pelos operadores do SAC de uma empresa de telecomunicação. Após o processamento dos dados textuais, foi possível extrair padrões que possibilitaram aos gestores antever a

ocorrência de ações estratégicas que contribuíram para a permanência do cliente no plano. Vasques et al., (2017) afirmam:

Empresas de telefonia podem se beneficiar das técnicas de Mineração de Textos, já que essas organizações possuem campos textuais presentes em alguns bancos de dados das centrais de atendimento telefônico ao cliente. Nesses campos, são relatadas as ocorrências efetuadas pelos clientes, as quais podem conter informações relevantes sobre suas experiências e preferências. Desse modo, a organização pode analisar os requisitos dos clientes e utilizá-los como critérios para a reavaliação do seu desempenho no mercado. Essa compreensão possibilitará o desenvolvimento de modelos de desempenho estratégico e de análises de lacunas processuais de negócios, fundamentais para a tomada de decisões.

Por fim, pode-se dizer que, está ocorrendo uma mudança na Gestão de relacionamento com os clientes, dado que essa nova abordagem procura a fidelização por intermédio do bom atendimento, contudo, a empresa precisa analisar os anseios dos clientes e utilizá-los como parâmetros para a avaliação de desempenho de seus processos internos. Entretanto, esses modelos são influenciados e orientados pelas informações dos clientes, que primeiramente devem ser extraídas e analisadas. Dessa maneira, a organização poderá comparar a diferença entre seu desempenho real e o desempenho ideal, pois, as empresas estão em busca de estratégias que deem respostas aos anseios de um mercado cada vez mais competitivo (VASQUES ET AL; 2017).

3 MÉTODO

Para o desenvolvimento desta pesquisa, fez-se necessário a utilização de metodologia adequada para alcançar os objetivos propostos. Sendo assim, este capítulo, tem por objetivo esclarecer os procedimentos, definições e escolhas adotados para sua realização, em consonância aos preceitos do processo formal e sistemático do método científico e do processo de pesquisa, bem como seu planejamento e técnicas aplicadas.

3.1 CARACTERIZAÇÃO DA PESQUISA

A pesquisa desenvolvida nesse estudo se classifica, segundo GIL (2002) como exploratória e descritiva. Exploratórias, são aquelas pesquisas que tem com objetivo incrementar a familiaridade com o tema, para torná-lo mais explícito, ou para construir hipóteses e descritiva porque, também segundo GIL (2002), pesquisas descritivas pretendem estabelecer relações entre variáveis e descrever características de uma população ou fenômeno.

Segundo Meirelles (2014) as metodologias quantitativas e, em especial, a estatística, permitem ao pesquisador que dela faça uso por meio da manipulação de uma quantidade infindável de dados verificar tendências, perceber padrões de comportamento e ação, prever resultados, analisar prevalências e riscos, definir estratégias de ação e elaborar modelos de análise. Tudo isto, a um custo relativamente baixo, de forma rápida e com o uso de aplicações informáticas específicas para as Ciências Sociais, as quais facilitam a elaboração de rotinas e a manipulação de grandes quantidades de dados.

3.2 DESCRIÇÃO DA BASE DE DADOS

Os dados utilizados no presente trabalho, foram cedidos pela Empresa Tri Telecom, uma empresa nacional do ramo de telecomunicações, e são provenientes

dos sistemas operacionais SOFIA e SENIOR. Os dados foram filtrados por clientes do tipo residencial, ou seja, clientes “pessoa física” dado que a empresa também possui clientes que são “pessoa jurídica”, os quais não foram objeto do estudo. A retirada da informação desta subpopulação de clientes originou um conjunto de dados com 10 868 registros, todos clientes “pessoa física” e com contrato do tipo residencial. Dos dados disponíveis, foram extraídos vários indicadores de diferentes categorias de informação: demográfica, socioeconômica, específica do negócio e relativas ao relacionamento entre a empresa e o cliente.

Relativamente ao intervalo de tempo dos dados extraídos, foi considerada a informação dos clientes, correspondente aos últimos 2 anos, visto que o objetivo é construir um modelo preditivo, capaz de a partir dos dados operacionais prever o abandono dos clientes, contudo é necessário considerarmos os clientes ativos.

Posteriormente, dentre o conjunto de variáveis à disposição nos dados extraídos, foi feita uma seleção daquelas que se considerou possuírem algum valor preditivo relativamente ao possível abandono de um cliente. Foi realizada uma limpeza e transformação dos dados escolhidos, e foi realizado o tratamento dos *outliers* e de valores nulos, que prejudicavam a qualidade dos dados. Para além destas operações, foi ainda realizada a categorização de algumas variáveis contínuas, com o intuito de facilitar o posterior processo de modelação.

A empresa proveu acesso ao seu banco de dados, e estes foram exportados para arquivos no formato “xls” (Microsoft Excel) e posteriormente foram convertidos para o formato “csv” (comma-separated values), para poderem ser importados mais facilmente pelo programa de análise. Para fazer a análise desses dados e o treinamento dos modelos, utilizou-se o software IBM SPSS Statistics.

3.2.1 Tratamento dos dados

O tratamento estatístico dos dados e o desenvolvimento dos modelos preditivos foram realizados por meio do software SPSS - Statistical Packages for the Social Sciences (Pacotes Estatísticos para as Ciências Sociais), é um pacote de software analítico oferecido pela IBM que os pesquisadores utilizam para analisar

dados complexos de pesquisa. No mercado, existem muitas aplicações informáticas que cumprem essa função e o SPSS é apenas uma delas. Assim sendo, no presente estudo foi utilizado exclusivamente o software em específico, recorrendo de suas potencialidades e do modo como o SPSS pode ser utilizado com uma abordagem quantitativa com vistas ao estabelecimento de padrões e tendências de comportamento amostral relacionadas a uma população específica.

3.3 DEFINIÇÃO DAS VARIÁVEIS INDEPENDENTES

Para Agresti e Finlay (2017), a estratégia de seleção das variáveis explicativas é um fator crítico para a efetividade de um modelo. Pode parecer óbvio incluir todo preditor potencialmente útil, porém, infelizmente, de modo geral isso é inadequado, especialmente quando o número de preditores é grande. Dispor de variáveis extras que acrescentam pouco poder de previsão em um modelo, notadamente pela sobreposição de outras variáveis, gera desvantagens, tornando o modelo de mais difícil interpretação, o que pode resultar em erros padrão inflados das estimativas dos parâmetros, impossibilitando a avaliação das contribuições parciais das variáveis que são efetivamente importantes. Ainda, Agresti e Finlay (2017) salientam que é desejável que as variáveis explicativas estejam correlacionadas com a variável dependente, mas que apresentem baixa ou nenhuma correlação entre si.

Para a construção do modelo de regressão logística foram selecionadas doze variáveis independentes, além da *target*, que é o status do cliente (ativo ou churn). As variáveis escolhidas para a estruturação do modelo constam no Quadro 2.

Quadro 2 – Descrição das variáveis independentes

Variável	Descrição	Código	Escala
V1	Gênero	1 - Masculino	Discreta
		2 - Feminino	
V2	Idade	Em anos	Contínua
V3	Cidade	1 - Alvorada	Discreta
		2 - Viamão	
		3 - Cachoeirinha	
		4 - Gravataí	
		5 - Canoas	
		6 - Porto Alegre	
V4	Tempo na Base	Em Semanas	Contínua
V5	Tipo de Plano	1 - Fiber 50mega	Discreta
		2 - Fiber 100mega	
		3 - Fiber 200mega	
		4 - Fiber 300mega	
		5 - Fiber 400mega	
		6 - Fiber 500mega	
		7 - Fiber 600mega	
V6	Telefonia	1 - Sim	Discreta
		0 - Não	
V7	Ponto Extra	1 - Sim	Discreta
		2 - Não	
V8	Atividade corretiva	Número de ocorrências	Discreta
V9	Manutenção Preventiva	Número de ocorrências	Discreta
V10	Manutenção Residencial	Número de ocorrências	Discreta
V11	Troca de Flat	Número de ocorrências	Discreta
V12	Incidência de cobrança	Número de ocorrências	Discreta
V13	Churn	1 - Sim	Discreta
		0 - Não	

Fonte: elaborado pelo autor

3.3.1 Modelos Preditivos

Análise preditiva é o ramo da mineração de dados objetivo estimar as probabilidades de que eventos ocorrerem, bem como prever as tendências. Para cada interação do cliente com a empresa, há um conjunto de ações que possuem um registro histórico. Com isso é possível agrupar os dados em grupos similares para produzir uma classificação. O agrupamento de dados, é um termo usado para a utilização de métodos numéricos de análise de dados com o objetivo de descobrir grupos homogêneos de objetos. Os sistemas *business intelligence* (BI) ajudavam os executivos a entender o que aconteceu, mas havia a necessidade de ferramentas que predissesse o que aconteceria e para onde o seu negócio estava indo, para preencher essa lacuna que surgiram os modelos preditivos.

3.3.2 Métodos de modelos preditivos

Conforme demonstrado no quadro 3, são diversas as técnicas passíveis de serem implementadas para execução de modelos preditivos de *churn*, porém algumas se destacam por seus resultados, notadamente as Redes Neurais Artificiais, Florestas Aleatórias (Random Forests) e Regressão Logística, sendo esta a escolhida para execução dos trabalhos. A definição da escolha das técnicas de modelagem preditora a ser utilizada é um *trade-off* entre a precisão e a interpretabilidade que se pretende obter, conforme Veloso (2013). Neslin et al. (2006) ponderam que técnicas de mais fácil interpretação podem ter uma performance preditiva não tão boa. Modelos como Random Forests e Redes Neurais Artificiais podem, por vezes, apresentar melhores resultados nas análises preditivas, mas são mais sensíveis a base de dados. Outras técnicas consideradas de mais fácil interpretação, como Regressão Logística, permitem interpretar as razões e sintomas do *churn*.

Quadro 3 - Métodos de modelos preditivos

Método	Definição
Redes Neurais	A utilização de dados na criação de bases de conhecimento é o conceito-chave das redes neurais. Em vez de trabalhar com regras explícitas como os sistemas especialistas, as redes neurais utilizam critérios mais complexos e implícitos, baseados no aprendizado a partir de exemplos (TSAI; LU, 2009).
Support Vector Machine	Método muito efetivo para reconhecimento de padrões de propósito geral. Intuitivamente, o SVM aprende o limite entre as amostras pertencentes a duas classes, mapeando as amostras de entrada em um espaço de alta dimensão e buscando um hiperplano de separação. As máquinas de vetores de suporte (SVMs) usam um modelo linear para implementar limites de classes não-lineares através de vetores de entrada de mapeamento não-lineares em um espaço de grande dimensão (SHMILOVICI, 2005).
Regressão Logística	Relaciona um conjunto de variáveis independentes com uma variável dependente categórica (GUANGLI et al., 2011).
Redes Bayesianas	Rede Bayesiana é uma representação compacta de uma tabela de conjunção de probabilidades. Modelo gráfico que representa de forma simples as relações de causalidade das variáveis de um sistema (MENÊZES; FIRMINO DROGUETT, 2005).
Árvore de Decisão	Com base nos registros do conjunto de treinamento, uma árvore é montada; a partir dela, pode-se classificar a amostra desconhecida sem necessariamente testar todos os valores dos seus atributos. Toda informação sobre cada objeto (caso) a ser classificado deve poder ser expressa em termos de uma coleção fixa de propriedades ou atributos. (YU et al., 2010)
Randon Forest	Combinação de preditores de árvores de modo que cada árvore depende dos valores de um vetor aleatório amostrado independentemente e com a mesma distribuição para todas as árvores na floresta (BREIMAN, 2001).
Análise de Cluster	Na análise de agrupamentos (<i>cluster analysis</i>) a similaridade entre duas amostras pode ser expressa como uma função da distância entre os dois pontos representativos destas amostras no espaço n- dimensional. Quanto menor a distância entre os pontos, maior semelhança entre as amostras. (QIAN; JIANG; TSUI, 2006)

Fonte: Elaborado pelo autor adaptada de SILVA (2017)

3.3.3 Regressão Logística

O modelo de Regressão Logística (RL) é uma técnica de análise multivariada cuja variável dependente (ou de resposta) é uma variável binária. Esta técnica permite ainda estimar a probabilidade de ocorrência de um evento (no presente caso, o evento de um cliente abandonar) e identificar as variáveis independentes que contribuem efetivamente para a predição da variável de resposta. Mesmo quando a resposta de Interesse (variável dependente) não é originalmente do tipo binário, alguns investigadores tendem a torná-la dicotômica de modo que a probabilidade de sucesso possa ser modelada através da regressão logística. No caso de um modelo de previsão de *churn* – abandono = sim ou não – é modelada a proporção de resposta de uma das duas classes em função das variáveis independentes. Logo, a variável resposta de interesse Y pode ser classificada como igual a 1, geralmente denotado para a resposta de maior interesse, também chamado de “sucesso”, que pode ser utilizado no caso de um cliente ser *churner*, ou igual a 0, caso contrário.

Sharma (1996); Hair et al., (1998); Morgan e Griego, (1998) enquadram esta técnica na classe de métodos estatísticos multivariados de dependência, pois relaciona um conjunto de variáveis independentes com uma variável dependente categórica. De forma geral, as técnicas de discriminação procuram encontrar uma função ou conjunto de funções que discrimine os grupos definidos pela variável categórica, visando a minimizar erros de classificação. Segundo Hosmer e Lemeshow (1989), a técnica de regressão logística tornou-se um método padrão de análise de regressão para variáveis medidas de forma dicotômica. O mesmo modelo pode ser utilizado com enfoque discriminatório, conforme descrevem Krzanowski (1988) e McLachlan (1992). Esses autores argumentam que o modelo logístico de discriminação pode ser utilizado de forma bem mais geral, pois não faz suposições quanto à forma funcional das variáveis independentes, e o número de parâmetros envolvidos no processo de estimação provavelmente será menor. (Quadro 4).

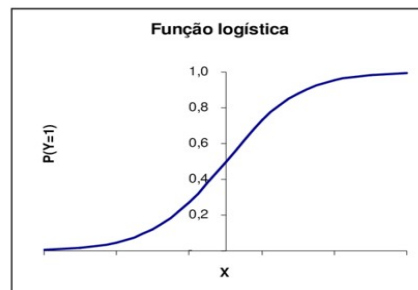
Quadro 4 - Principais Características da regressão logística

Principais Características da Regressão Logística	
A modelagem é bem conhecida, conceitualmente simples, frequentemente usada em marketing, especialmente no nível do consumidor individual;	Não é necessário supor normalidade multivariada.
A facilidade de interpretação do <i>logit</i> ;	É uma técnica mais genérica e mais robusta, pois sua aplicação é apropriada em grande variedade de situações.
A modelagem <i>logit</i> mostrou fornecer bons e robustos resultados em estudos gerais preditivos	é uma técnica similar à regressão linear múltipla.
Mais especificamente nas abordagens de métricas do marketing, foi demonstrado por vários autores que a modelagem <i>logit</i> pode até mesmo superar métodos mais sofisticado	Na regressão logística, a probabilidade de ocorrência de um evento pode ser estimada diretamente.
Enquanto a análise de sobrevivência nos daria o tempo estimado para o evento (neste caso, <i>churn</i>), estamos interessados apenas na probabilidade estimada do evento dentro de um dado período.	
Conforme Van Den Poel (2005)	Conforme Hair et al. (1998)

Elaborado pelo autor

No caso da variável dependente Y assumir apenas dois possíveis estados (1 0) e haver um conjunto de p variáveis independentes X_1, X_2, \dots, X_p , no modelo de regressão logística os coeficientes B_0, B_1, \dots, B_p são estimados a partir do conjunto de dados, pelo método da máxima verossimilhança, que encontra uma combinação de coeficientes que maximiza a probabilidade de a amostra ter sido observada (Hosmer e Lemeshow, 1989). Observa-se que a curva logística tem comportamento probabilístico no formato da letra S, o que é característica da regressão logística. (Figura 2).

Figura 2 – Representação de uma Função logística



Elaborada pelo autor

Como se pode observar a função logística pode receber como *input* qualquer valor desde menos infinito até mais infinito, devolvendo sempre como resultado um valor, correspondendo a uma probabilidade entre 0 e 1 (Agresti, 2017). Relativamente à cardinalidade que a variável dependente pode assumir, as regressões logísticas são divididas em dois subtipos: regressões logísticas binomiais e regressões logísticas multinomiais. As regressões logísticas binomiais são usadas quando a variável de resposta, ou variável dependente, assume apenas dois valores, normalmente 1 ou 0, indicando verdadeiro/sucesso ou falso/insucesso respectivamente (Weisberg, 2005). Apesar de ser um método relativamente simples, a regressão logística pode produzir resultados satisfatórios e, por isso, é amplamente utilizada em diversas aplicações na indústria (BATISTA, 2019).

3.4 AVALIAÇÃO DO MODELO PREDITIVO

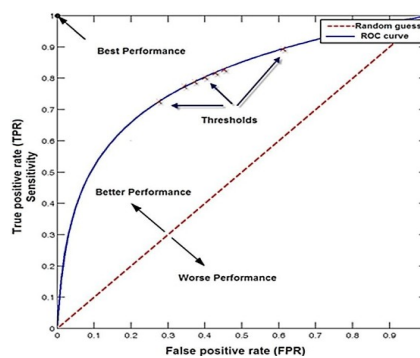
A validação tem como objetivo verificar a acurácia e precisão do método escolhido para a construção do modelo. Para a validação do modelo foram usados a curva ROC (Receiver Operating Characteristic), a Matriz de Confusão, além do teste de colinearidade.

3.4.1 Curva ROC (Receiver Operating Characteristic)

Segundo KARAHOCA, A.; KARAHOCA, D.; AYDIN, N. GSM a curva ROC

(Receiver Operating Characteristic, é uma representação das relações entre a taxa de verdadeiros positivos (ou benefícios) e a taxa de falsos positivos (ou custos), desenhada nos eixos x e y em uma escala linear (KARAHOCA, A.; KARAHOCA, D.; AYDIN, N. GSM). A ROC representa as relações entre a taxa de *Customer Churn* prevista corretamente como *Customer Churn* e a taxa de não-*churners* prevista incorretamente como *Customer Churn*. Em outras palavras, a ROC fornece implicações relativas entre benefícios e custos. A curva ROC consiste de pontos correspondentes aos resultados da previsão. A Figura 3 apresenta um exemplo de uma curva ROC (HASSOUNA et al., 2015).

Figura 3 – Exemplo de uma Curva ROC



Fonte: elaborada pelo autor

3.4.2 Matriz de Confusão

Matriz de Confusão (Figura 4) apresenta a quantidade de casos preditos como sendo da classe 0 e que realmente são da classe 0 (chamados de verdadeiro negativo), a quantidade de casos preditos como sendo da classe 0, mas que são na verdade da classe 1 (falso negativo), a quantidade de casos preditos como sendo da classe 1 e que realmente são da classe 1 (verdadeiro positivo), e os casos preditos como sendo da classe 1, mas que na verdade são da classe 0 (falso positivo).

Figura 4 – Exemplo de matriz de confusão

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Fonte: Elaborado pelo autor

- Verdadeiro Positivo (true positive - TP) - ocorre quando, no conjunto real, a classe que estamos buscando foi prevista corretamente. Por exemplo, quando houve um *Customer Churn* e o modelo previu corretamente que haveria.
- Falso Positivo (false positive - FP) - ocorre quando no conjunto real, a classe que estamos buscando foi prevista incorretamente. Por exemplo, não houve um *Customer churn*, mas o modelo previu que haveria . Corresponde ao Erro tipo I.
- Falso negativo (false negative - FN) - ocorre quando, no conjunto real, a classe que não estamos buscando foi prevista incorretamente. Por exemplo, quando houve um *Customer Chrun* e o modelo previu incorretamente que não haveria. corresponde ao Erro tipo II.
- Verdadeiro Negativo(true negative - TN) - ocorre quando, no conjunto real, a classe que não estamos buscando foi prevista corretamente. Por exemplo, não houve um *Customer Chrun*, e o modelo previu corretamente que não haveria.

4 ANÁLISE DOS RESULTADOS

As variáveis explicativas referem-se às informações constantes no banco de dados dos clientes, relativo aos clientes que estavam ativos ou ficaram inativos entre Jan 2018 a Dez 2020. Classificou-se os clientes como ativos (cliente com contrato ativo e

inativo (cancelou ou teve o contrato cancelado nesse período).

4.1 VERIFICAÇÃO DE MULTICOLINEARIDADE

Conforme Cerqueira (2016), a multicolinearidade é um problema computacional que se desenvolve quando duas ou mais variáveis independentes possuem forte correlação. O resultado é uma Matriz de Correlação com variabilidade única (próximo de zero) tendo em vista o efeito da divisão de um número por uma variação absoluta extremamente pequena (próxima de zero). O resultado desta divisão é um número, com um número de casas decimais bastante elevado, o que torna a aproximação computacional totalmente ineficaz. A existência de multicolinearidade introduz erros grosseiros no resultado da regressão, produzindo sérios erros na previsão da variável dependente. Felizmente a multicolinearidade é de fácil detecção e correção. A ocorrência de multicolinearidade dá-se quando um ou mais dos testes t assume resultado(s) muito pequenos (não significativos) e os valores de r^2 ou da estatística F são muito grandes. Se este for o caso, a Matriz de Correlação deve ser avaliada e, possivelmente, uma das Variáveis Independentes com forte correlação ser eliminada (geralmente, um coeficiente de correlação superior ou próximo a 0,7 entre duas Variáveis Independentes indica problemas de multicolinearidade). A multicolinearidade sendo um problema no ajuste do modelo pode causar impactos na estimativa dos parâmetros. Podemos diagnosticar Multicolinearidade por meio do VIF (Variance Inflation Factor).(Tabela 1).

Tabela 1 - Estatísticas de colinearidade

Coefficientes^a

Modelo	Coefficients não padronizados		Coefficients padronizados	t	Sig.	Estatísticas de colinearidade	
	B	Erro Padrão	Beta			Tolerância	VIF
1 (Constante)	1,459	,018		79,266	,000		
Gênero	-,012	,008	-,012	-1,493	,135	,999	1,001
Incidência_cobrança	,025	,013	,016	1,984	,047	,974	1,026
Troca_de_flat	-,014	,015	-,008	-,950	,342	,985	1,016
Manutenção_Preventiva	-,074	,018	-,034	-4,221	,000	,968	1,033
Atividade_corretiva	-,058	,065	-,007	-,898	,369	,994	1,006
Telefonia	-,003	,008	-,003	-,390	,696	,999	1,001
Ponto_extra	-,064	,011	-,048	-5,828	,000	,927	1,079
tempo_na_base	-,024	,000	-,478	-50,709	,000	,706	1,417
Tipo_plano	-,181	,004	-,507	-47,887	,000	,560	1,786
idade	-,001	,000	-,027	-3,389	,001	,998	1,002
Cidade	-,062	,002	-,256	-26,587	,000	,676	1,478

a. Variável Dependente: churn

Fonte: Output SPSS

Se o VIF for igual a 1 não há multicolinearidade entre os fatores, mas se o VIF for maior que 1, as preditoras podem estar moderadamente correlacionadas. Um VIF entre 5 e 10 indica alta correlação, o que pode ser problemático. E se o VIF for acima de 10, você pode assumir que os coeficientes de regressão estão mal estimados devido à multicolinearidade. Quando os valores de tolerância forem superiores a 0,1 também indicativo de não incidência de multicolinearidade.

4.2 ÍNDICE DE EXATIDÃO DO TESTE

Tabela 2 – Área sob a curva
Área sob a curva

Variável(eis) de resultado de teste: Tipo Plano

Área	Erro Padrão ^a	Sig. assintótico ^b	Intervalo de Confiança 95% Assintótico	
			Limite inferior	Limite superior
,728	,005	,000	,719	,738

A variável ou variáveis de resultado de teste: Tipo Plano possuem pelo menos um vínculo entre o grupo de estado real positivo e o grupo de estado real negativo. As estatísticas podem ser enviesadas.

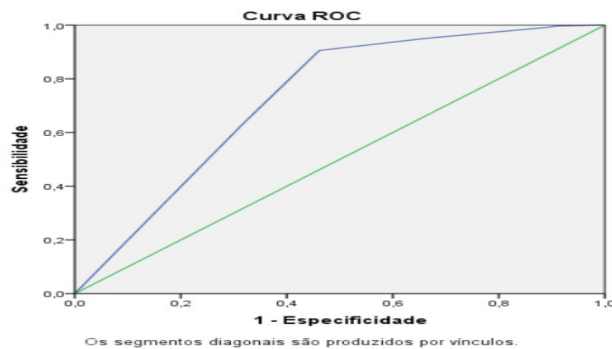
a. Sob a suposição não paramétrica

b. Hipótese nula: área real = 0,5

Fonte: Output SPSS

Um teste totalmente incapaz de discriminar indivíduos *churn* e não *churn*, teria uma área sob a curva de 0,5 (Seria a hipótese nula). Acima de 0,70 é considerado desempenho satisfatório. O cálculo da área sob curva do presente estudo foi de 0,728.(Tabela 2),(Figura 9).

Figura 5 – Curva ROC



Fonte: Output SPSS

4.3 ÍNDICE DE AJUSTE DO MODELO (MATRIZ DE CONFUSÃO)

O Software denomina a matriz de confusão como tabela de classificação (tabela 3).

O modelo conseguiu classificar acertadamente em 77,7 % os casos classificados como não *churn*, em 82,8 % os casos classificados como *churn*. E apresentou uma excelente acurácia global de 80,5 %.

Tabela 3 -Tabela de Classificação

Tabela de Classificação^a

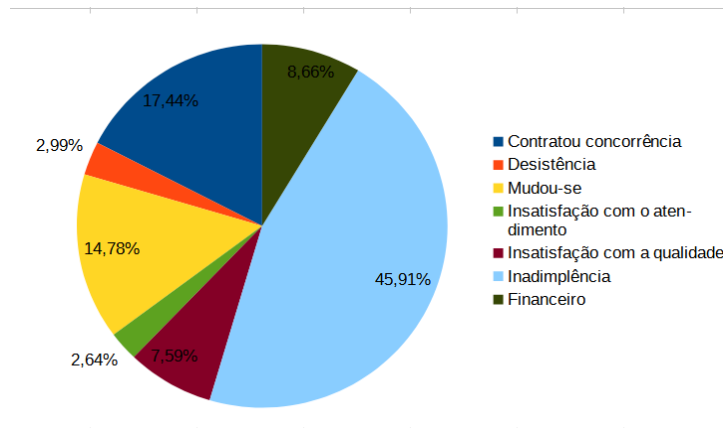
Observado		Previsto		
		churn		Porcentagem correta
		não	sim	
Etapa 1	churn não	3899	1118	77,7
	sim	1004	4846	82,8
Porcentagem global				80,5

a. O valor de recorte é ,500

Output SPSS

Dentre as informações extraídas, e para posterior comparação e análise, foi efetuado um levantamento dos clientes *churn* no período de janeiro/20 a dezembro/20.(Gráfico 4).

Gráfico 4 - Clientes cancelados 2020



4.4 INTERPRETAÇÃO DOS RESULTADOS DA REGRESSÃO LOGÍSTICA

As primeiras tabelas descrevem alguns aspectos da configuração das variáveis utilizadas na análise, no presente estudo foram incluídos na análise 10 867 casos.(Tabela 4).

Tabela 4 – Resumo do processamento do caso

Resumo de processamento do caso

Casos não ponderados ^a		N	Porcentagem
Casos selecionados	Incluído na análise	10867	100,0
	Casos ausentes	1	,0
	Total	10868	100,0
Casos não selecionados		0	,0
Total		10868	100,0

a. Se a ponderação estiver em vigor, veja a tabela de classificação para o número total de casos.

Fonte: Output SPSS

4.4.1 Bloco 1

As três próximas tabelas sob o título principal, "Bloco 0: Bloco inicial", mostram a situação em que nenhuma das variáveis independentes foram adicionados ao modelo, ou seja, o modelo inclui apenas a constante. O interesse nesta informação é apenas como uma comparação com o modelo com todas as variáveis independentes acrescentadas.(Tabela 5).

Tabela 5 – Bloco Inicial Bloco 0

Bloco 0: Bloco InicialTabela de Classificação^{a,b}

Observado			Previsto		
			churn		Porcentagem correta
			não	sim	
Etapa 0	churn	não	0	5017	,0
		sim	0	5850	100,0
Porcentagem global					53,8

a. A constante está incluída no modelo.

b. O valor de recorte é ,500

Variáveis na equação

	B	S.E.	Wald	df	Sig.	Exp(B)
Etapa 0 Constante	,154	,019	63,727	1	,000	1,166

Variáveis não presentes na equação^a

			Pontuação	df	Sig.
Etapa 0	Variáveis	Gênero(1)	,215	1	,643
		Telefonia	,193	1	,660
		Ponto_extra	161,339	1	,000
		Incidência_cobrança	13,185	1	,000
		Troca_de_flat	13,563	1	,000
		Manutenção_Preventiva	159,036	1	,000
		Manutençãoresidencial	13,185	1	,000
		Atividade_corretiva	7,599	1	,006
		idade	15,359	1	,000
		tempo_na_base	293,246	1	,000
		Cidade	1888,678	1	,000
		Tipo_plano	2153,361	1	,000

Fonte: Output SPSS

A “Tabela de Classificação”, mostra que, sem quaisquer variáveis independentes, o “melhor palpite” é simplesmente assumir que todos os clientes são *churn* e nesse caso acerta em 53,8 % dos casos.

Na tabela “variáveis não presentes na equação” é possível verificar que com exceção da variável categórica (gênero) e da variável (telefonia) as demais são estatisticamente significativas ($p < 0,0005$), e devem participar do modelo.

4.4.2 Testes Bloco 1

Nesse bloco o SPSS fornece um conjunto de testes que podem ser utilizados para verificar a qualidade do modelo (Omnibus Test) os atributos e desempenho das variáveis independentes (Cox & Snell R2, Nagelkerke R2) e do modelo Hosmer and Lemeshow test.

Tabela 6 – Bloco 1 Método Enter

Bloco 1: Método = Enter

Testes de coeficientes de modelo Omnibus

		Qui-quadrado	df	Sig.
Etapa 1	Etapa	5631,503	11	,000
	Bloco	5631,503	11	,000
	Modelo	5631,503	11	,000

Resumo do modelo

Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	9369,442 ^a	,404	,540

a. Estimação finalizada no número de iteração 5 porque as estimativas de parâmetro mudaram foram alteradas para menos de ,001.

Teste de Hosmer e Lemeshow

Etapa	Qui-quadrado	df	Sig.
1	178,716	8	,000

Fonte: Output SPSS

Testes de Coeficiente de modelos Omnibus: O objetivo destes testes é verificar a qualidade do modelo e tem um qui quadrado como hipóteses:

H0: ajuste do modelo atual = ajuste do modelo sem previsores.

H1: ajuste do modelo atual \neq ajuste do modelo sem previsores.

Ou seja, a inclusão das variáveis não melhorou a previsão, nesse caso como $p < 0,05$, devemos considerar a hipótese alternativa. Concluímos que o modelo com as variáveis independentes é melhor que o modelo sem nenhum previsor.

Cox-Snell R2 Este teste é comparável ao R-quadrado da regressão linear. Ou seja, indica que 40 % das variações ocorridas no Log da razão de chances são explicadas pelo conjunto das variáveis independentes. Trata-se de um mecanismo que pode ser utilizado para comparar o desempenho de modelos concorrentes. Entre duas equações logísticas igualmente válidas. Deve-se preferir o que apresente o Cox-Snell R2 mais elevado.

Nagelkerke R2 sua finalidade é a mesma do Cox-Snell R2 Na prática a única diferença está em se fazer mais compreensível que o Cox-Snell R2 O Nagelkerke R2 é uma versão do Cox - Snell adaptada para fornecer resultados entre 0 e 1. Conclui-se que o modelo é capaz de explicar cerca de 54,% das variações registradas na variável dependente.

Hosmer e Lemeshow no quadro “Teste de Hosmer e Lemeshow” para avaliar a significância estatística do teste. O valor de p deve ser $\geq 0,05$, é encontrado na tabela de título similar. Para este modelo, o teste de Hosmer e Lemeshow é estatisticamente significativo ($p = 0,00$), o que indica que o modelo não consegue responder adequadamente o comportamento da variável dependente. A tabela 7 demonstra aonde se encontra essa falha.

Tabela 7 - Tabela de contingência para teste de Hosmer e Lemeshow

Tabela de contingência para teste de Hosmer e Lemeshow

		churn = não		churn = sim		Total
		Observado	Esperado	Observado	Esperado	
Etapa 1	1	990	1044,104	97	42,896	1087
	2	913	956,493	174	130,507	1087
	3	913	860,301	174	226,699	1087
	4	777	730,502	310	356,498	1087
	5	603	547,854	484	539,146	1087
	6	417	377,589	670	709,411	1087
	7	233	241,817	854	845,183	1087
	8	102	143,093	985	943,907	1087
	9	29	79,553	1058	1007,447	1087
	10	40	35,695	1044	1048,305	1084

Output SPSS

Este teste mede o grau de acurácia do modelo logístico, este indicador corresponde a um teste do qui-quadrado que consiste em dividir o número de observações em cercas de dez classes e, em seguida, comparar as frequências preditas com as observadas. A finalidade desse teste é verificar se existem diferenças significativas entre as classificações realizadas pelo modelo e a realidade observada. Busca-se não rejeitar a hipótese de que não existem diferenças entre os valores preditos e observados.

H0: não há diferenças significativas entre os resultados os preditos pelo modelo e os observados.

H1: há diferenças significativas entre os resultados os preditos pelo modelo e os observados.

Variáveis na equação Na tabela “variáveis na equação” (Tabela 8) são apresentados os parâmetros da equação e sua estatística para cada previsor e para a constante, onde B representa o coeficiente de cada previsor que modifica a saída resultante da mudança de uma unidade na variável previsor. É também apresentada a estatística Wald e seu respectivo valor de “p” (Sig.) que informa se o coeficiente “b” para dado previsor é significativamente diferente de zero. Na coluna Exp(B) é descrita a taxa de mudança na chance que resulta do aumento de uma

unidade do previsor. Igualmente é fornecido o intervalo de confiança de 95 % para EXP (B) e quando este intervalo conter o valor 1 evidência que essa variável não contribui significativamente no aumento de chance do evento ocorrer.

Tabela 8 – Variáveis na equação

		Variáveis na equação						95% C.I. para EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Inferior	Superior
Etapa 1 ^a	Gênero	-,068	,052	1,720	1	,190	,935	,845	1,034
	Ponto_extra	-,308	,077	16,062	1	,000	,735	,632	,854
	Incidência_cobrança	,169	,095	3,160	1	,075	1,184	,983	1,426
	Atividade_corretiva	-,443	,465	,909	1	,340	,642	,258	1,597
	Troca_flat	-,062	,091	,465	1	,495	,940	,787	1,123
	Telefonia	-,027	,052	,261	1	,610	,974	,880	1,078
	Manutenção_Preventiva	-,710	,145	23,983	1	,000	,492	,370	,653
	idade	-,008	,002	16,950	1	,000	,992	,988	,996
	Cidade	-,350	,015	561,371	1	,000	,705	,685	,725
	tempo_na_base	-,147	,004	1703,479	1	,000	,863	,857	,869
	Tipo_plano	-1,223	,031	1568,512	1	,000	,294	,277	,313
	Constante	6,169	,147	1752,210	1	,000	477,547		

a. Variável(is) inserida(s) na etapa 1: Gênero, Ponto_extra, Incidência_cobrança, Atividade_corretiva, Troca_flat, Telefonia, Manutenção_Preventiva, idade, Cidade, tempo_na_base, Tipo_plano.

Output SPSS

A partir destes resultados é possível verificar que as variáveis Gênero, incidência de_cobrança, atividade_atividade corretiva, troca-flat e telefonia, não contribuem significativamente para a ocorrência ou não do *churn*, pois o intervalo de confiança para o EXP (B) contém o valor 1.

A variável ponto_extra representa um fator de proteção para ocorrência de churn, pois o valor de “B” é -0,308 e o valor de EXP (B) é 0,735, significando que tudo mais constante, ao aumento de uma unidade no previsor aumenta em 0,735 % a chance do cliente não cancelar. Igualmente as variáveis, manutenção_preventiva e tempo_na base contribuem para que o *churn* não ocorra, pois representam 0,482 % e 0,863 %, respectivamente de chance do evento não ocorrer sempre que houver o incremento de uma unidade no previsor. *ceteris paribus*.

Idade, e o tipo_de-plano representam um fator que influência para que o *churn* não ocorra, pois o incremento de um ano na idade do cliente diminui em 0,988 a chance do *churn* ocorrer. A depender do Tipo_de_plano existe uma *odds ratio* de

0,294 do cliente não abandonar a empresa.

O SPSS disponibiliza a medida de associação razão de prevalência (RP) que uma medida de associação que visa mensurar a relação de um desfecho binário com as variáveis em estudo. Conforme a tabela 9 é possível observar que dos clientes que possuem tipo de plano Fiber 50 mega 69,9 % cancelaram e que a medida que os planos tem um acréscimo no valor os percentuais diminuem.

Tabela 9 Tipo-Plano X Churn

Tipo_plano * churn Tabulação cruzada

			churn		Total
			não	sim	
Tipo_plano	Fiber 50mega	Contagem	1594	3703	5297
		% em Tipo_plano	30,1%	69,9%	100,0%
	Fiber 100mega	Contagem	721	1595	2316
		% em Tipo_plano	31,1%	68,9%	100,0%
	Fiber 200mega	Contagem	995	260	1255
		% em Tipo_plano	79,3%	20,7%	100,0%
	Fiber 300mega	Contagem	1166	244	1410
		% em Tipo_plano	82,7%	17,3%	100,0%
	Fiber 400mega	Contagem	89	29	118
		% em Tipo_plano	75,4%	24,6%	100,0%
	Fiber 500mega	Contagem	399	19	418
		% em Tipo_plano	95,5%	4,5%	100,0%
	Fiber 600mega	Contagem	53	0	53
		% em Tipo_plano	100,0%	0,0%	100,0%
Total		Contagem	5017	5850	10867
		% em Tipo_plano	46,2%	53,8%	100,0%

Output SPSS

Na associação entre os clientes que cancelaram e a incidência_cobrança é possível verificar que dos clientes que tiveram incidência de cobrança, em 59,2 % dos casos ocorreu o *churn*. (Tabela 10).

TABELA 10 INCIDÊNCIA_COBRANÇA x CHURN

Incidência_cobrança * churn Tabulação cruzada

			churn		Total
			não	sim	
Incidência_cobrança	0	Contagem	4596	5238	9834
		% em Incidência_cobrança	46,7%	53,3%	100,0%
	1	Contagem	422	612	1034
		% em Incidência_cobrança	40,8%	59,2%	100,0%
Total		Contagem	5018	5850	10868
		% em Incidência_cobrança	46,2%	53,8%	100,0%

Output SPSS

5 CONSIDERAÇÕES FINAIS

Com o aumento da competitividade entre as empresas, a retenção de clientes passou a ser uma atividade prioritária. Além disso, as campanhas de retenção revelam-se menos dispendiosas do que a aquisição de novos clientes e incidem sobre os clientes mais fiéis que eventualmente serão os mais rentáveis. A compreensão da necessidade de combater o *churn* de clientes leva as empresas a procurar conhecer os clientes e a forma de obter esse conhecimento é através do uso da base de dados obtida através da interação com os clientes no cotidiano da empresa. As informações relevantes sobre os clientes podem ser extraídas e tratadas recorrendo às ferramentas estatísticas ou de gerenciamento de *churn*. Uma das principais estratégias de retenção é a prevenção eficiente do *churn* de clientes, que visa, na sua essência, prever quais os clientes que estão em risco de abandonar a empresa voluntariamente, com o propósito de se agir de forma direcionada e antecipada de modo a evitar essa situação.

Este trabalho objetivava construir um modelo de previsão de *churn* de clientes, com a antecipação necessária de forma a ser possível agir a tempo de impedir que o *churn* ocorra e a perda de lucro associada a esse cliente. Em termos de objetivos este trabalho produziu um modelo que está de acordo com aquilo que tem sido publicado recentemente em termos de modelos preditivos de *churn*. Quanto ao objetivo específico, que propunha adquirir e adequar as bases de clientes o perfil dos clientes, encontrou-se no banco de dados da empresa Tri Telecom as informações necessárias. Tais achados estão em consonância com os referenciais teóricos que apontam que a gestão do relacionamento com o cliente (CRM) trata do gerenciamento cuidadoso de informações detalhadas sobre cada cliente e todos os pontos de contato com ele a fim de maximizar sua fidelidade.(Kotler e Keller,2006).

Há dois tipos de causas associadas à taxa de abandono: causas voluntárias e causas involuntárias (FERNANDES, A.P., 2007). Esse estudo apontou a inadimplência como a principal razão de *churn* (46%) e a atração dos concorrentes (17%) também é um elemento com grandes probabilidades de gerar deserção do cliente, observou-se ainda que o *churn* inevitável (mudou-se) respondeu por

aproximadamente (15%) dos casos de *churn*. Através dessa pesquisa bibliográfica constatou-se que a regressão logística tem sido extensivamente usada para estimar a ocorrência de *churn* em empresas de telecomunicação.

Foram utilizados dados reais da Empresa Tri Telecom e o passo inicial nesse sentido foi realizar a tarefa de compilar os dados, avaliar e escolher as variáveis para a partir daí dedicar-se na execução do modelo preditivo, para esse fim recorreu-se ao Software SPSS statistics, e o modelo de predição escolhido foi a regressão logística. Após o entendimento da aplicabilidade da equação estatística, foi realizada uma análise dos dados encontrados, em que se verificou que o modelo apresentou uma acurácia global de 80,5 %. Para o estudo foram definidas 12 variáveis independentes, além da constante; desse conjunto de preditoras 6 não contribuíram de maneira significativa para composição do modelo.

Dentre as limitações deste estudo, possivelmente a mais significativa refere-se ao padrão dos dados constantes do banco de dados da Tri Telecom, em virtude de alguns dados inconsistentes, dados *missing* e dados desatualizados. Mediante essa limitação, variáveis importantes deixaram de ser consideradas. Como complemento deste trabalho, seria importante acrescentar outros tipos de variáveis, relativas ao serviço a clientes, por exemplo, número de reclamações ou indicadores capazes de sinalizar o nível de satisfação dos clientes, tanto com a qualidade dos serviços, quanto ao atendimento. Estas variáveis poderiam ajudar a encontrar outros tipos de *churners*, e eventualmente constituir um modelo preditivo que apresente maior precisão.

A principal razão para a aplicação de técnicas de gerenciamento de *churn* numa campanha de fidelização de clientes é otimizar a aplicação dos recursos monetários disponíveis, e dessa maneira, evitar desperdiçar muito dinheiro em clientes que não estão, efetivamente, em risco de serem perdidos. A aplicação de técnicas de gerenciamento de *churn* permite diminuir o desperdício e aumentar o valor a investir sobre os clientes-alvo. É importante que se faça uma análise financeira prévia para avaliar se os proveitos da campanha justificam o investimento.

Com o modelo proposto apoiando uma boa estratégia de retenção, pode-se chegar a uma excelente redução de *churn* na empresa, contudo mesmo que as

análises preditivas permitam identificar os clientes que tendem ao fenômeno do *churn*, a empresa deve procurar atuar em outras frentes, como diminuir o CAC (custo de aquisição de clientes), investindo em estratégias de aquisição que permitam a escalabilidade e a previsibilidade no médio e longo prazo; aumentar o *lifetime value* (lucro líquido da vida do cliente), aplicando processos para manter o cliente por mais tempo e adotar uma cultura centrada no cliente (*Customer Success*), obtendo assim maior conhecimento sobre os clientes que estão em risco, pois abordagens em nível individual podem contribuir com elementos importantes para retenção dos clientes.

REFERÊNCIAS

AGRESTI, A.; FINLAY, B; **Métodos estatísticos para as ciências sociais**. 4. ed. Porto Alegre: Penso, 2017.

ASSUNÇÃO, Fernando, **Estratégias para tratamento de variáveis com dados faltantes durante o desenvolvimento de modelos preditivos**, dissertação de Mestrado em Ciências, USP, São Paulo, 2012.

BOGMANN, Itzhak Meir, **Marketing de relacionamento: estratégias de fidelização e suas implicações financeiras**. São Paulo: Nobel, 2002.

BARQUETTE, S., & Chaoubah, A. (2007). **Pesquisa de marketing** (2a ed.). São Paulo: Saraiva.

BERSON, Alex; SMITH, Stephen J.; THEARLING, Kurt. **Building data mining applications for CRM**. New York: McGraw-Hill, 2000. 510 p.

CASTELLS, Manuel. **A sociedade em rede – a era da informação**: economia, sociedade e cultura. Tradução de Roneide Venâncio Majer. 10. ed. São Paulo: Paz e terra, 2010. v. 1. 698 p.

CASTRO, L. N.; FERRARI, D. G.. **Introdução à mineração de dados**: conceitos básicos, algoritmos e aplicações. 1a Edição. São Paulo: Saraiva, 2016

CERQUEIRA, Rosa Ildebranda Pedrosa Bompastor; Mestrado em Bioestatística e Biometria, Universidade Aberta, 2016.

COSTA, José Alfredo Ferreira, **Uma aplicação de Mineração de dados no gerenciamento do churn em serviços de banda larga de telecomunicações**, IV Congresso Nacional de excelência em gestão . Niterói, 2008.

CISTER, A. M. **Mineração de dados para a análise de atrito em telefonia móvel**. Tese (Doutorado em Engenharia). Rio de Janeiro, 2005. Programa de Pós-Graduação em Engenharia Civil, COPPE-Civil. Universidade Federal do Rio de Janeiro.

D`ANGELO, Et Al. **Marketing de relacionamento junto a consumidores finais: um estudo exploratório com grandes empresas brasileiras**, Rev.adm.contemp.vol10 no1 Curitiba Jan./Mar.2006

DIEHL, Antônio Astor; TATIM, Denise Carvalho. **Pesquisa em Ciências Sociais Aplicadas: Métodos e Técnicas**. 1. ed. São Paulo: Pearson Prentice Hall, 2004.

Estudos em negócios IV / Margarida Gutierrez e Hélène Bertrand, organizadoras – Rio de Janeiro-Mauad, 2005

FRANCESCHI, Pietro Reinheier de, **MODELAGENS PREDITIVAS DE CHURN: O CASO DO BANCO DO BRASIL**; Dissertação de Mestrado, rograma de Pós-Graduação em Gestão e Negócios da Universidade do Vale do Rio dos Sinos – UNISINOS, 2019

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From **Data Mining to Knowledge Discovery in Databases**. *Artificial Intelligence Magazine*, v. 17, n. 3, p. 37-54, 1996a.

FERREIRA, Jorge B. **Mineração de dados na retenção de clientes em telefonia celular**. 2005. 93 f. Dissertação (Mestrado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica da Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2005.

FERNANDES NETO, André P.; MATTOZO, Teófilo C.; COSTA, José A. F. **Uma aplicação de mineração de dados no gerenciamento do churn em serviços de banda larga de telecomunicações**. In: CONGRESSO NACIONAL DE EXCELÊNCIA EM GESTÃO, 4., Niterói, 2008.

GALBREATH, J.; ROGERS, T. **Customer Relationship Leadership: A Leadership and Motivation Model for the XXI Century Business**. *TQM Magazine*, v.11, n.3, p.16-17, 1999.

GAUER, Jefferson José Cerutti. **Modelagem de evasão de clientes bancários adimplentes: identificação de padrões pelo histórico de suas operações**. / Jefferson José Cerutti Gauer– 2016.

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

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

HAIR JR., J.F.; WILLIAM, B.; BABIN, B.; ANDERSON, R.E. **Análise multivariada de dados**. 6.ed. Porto Alegre: Bookman, 2009

HENNIG-THURAU, T.; KLEE; A. **The impact of customer satisfaction and relationship quality on customer retention: a critical reassessment and model development**. *Psychology & Marketing*, v. 14, n. 8, p. 737-764, 1997.

HENNING,Thurau, T., Gwinner, K.P., Walsh, G., et al. (2004) **Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet**. *Journal of Interactive Marketing*, 18, 38-52.

HOSMER, D.W. and Lemeshow, S. (1989) **Applied Logistic Regression**. Willey & Sons, Hoboken, 373

HOFFMAN, K. D.; BATESON, J. E.G. **Essentials of Services Marketing**. Texas: The Dryden Press, 1997

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA

Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/industria/9300-contas-nacionais-trimestrais.html?>>. Acesso em 19/09/2020.

KARAOCA, A.; KARAOCA, D.; AYDIN, N. GSM churn management using an adaptive neuro-fuzzy inference system. In: THE 2007 INTERNATIONAL CONFERENCE ON INTELLIGENT PERVASIVE COMPUTING, Valencia, Spain, 2007. [Proceedings], p. 323- 326.

KOTLER, Philip; ARMSTRONG, Gary. **Princípios de Marketing**, São Paulo: Atlas, 1999

KUMAR V., REINARTZ W. **Impact of CRM on Marketing Channels**. In: **Customer Relationship Management**. Springer Texts in Business and Economics. Springer, Berlin, Heidelberg, 2012

LIGHT, Ben. CRM Packaged Software: **A Study of Organizational Experiences**. Journal of Business Process Management, v.9, n.5, p.603-616, 2003.

LINOFF, Gordon S.; BERRY, Michael J. A. **Data mining techniques**: for marketing, sales, and customer relationship management. 3. ed. New York: John Willey & Sons, 2011. 829 p.

LUIZON, Bruna Franciany Girata Luizon, **Análise preditiva de churn em um e-comerc** Monografia.Curitiba,2019 Universidade Federal do Paraná

MALHOTRA, Naresh K. **Pesquisa de Marketing**: uma orientação aplicada. 3. ed. Porto Alegre: Bookman, 2001.

MARCONI, M. A; LAKATOS, E. M. **Técnicas de pesquisa**: planejamento e execução de pesquisas, amostragens e técnicas de pesquisas, elaboração e interpretação de dados. 3ª ed. São Paulo: Atlas, 1996.

MATTISON, Rob, The Telco Churn Management Handbook, USA-2005.

MCKENNA, R. **Marketing de relacionamento**. Rio de Janeiro: Campus, 1999.

MEIRELLES, Mauro **O Uso do SPSS (Statistical Package for the Social Sciences) na Ciência Política: uma breve introdução**, Pelotas 2014

MILAN, Gabriel Sperandio, **A prática do marketing de relacionamento e a retenção de clientes: um estudo aplicado em um ambiente de serviços** Tese (Doutorado em Engenharia). Porto Alegre, 2006. Programa de Pós-Graduação em Engenharia de Produção, UFRGS. Universidade Federal do Rio Grande do Sul.

MOZER, M. C; WOLNIEWICZ, R; GRIMES, D. B; JOHNSON, E; KAUSHANSKY, H. Churn reduction in the wireless industry. In: SOLLA S. A; LEEN T. K; MUELLER K. R (Eds) **Advances in neural information processing systems**. Cambridge, MA: MIT Press, 2000. p. 935-941.

NESLIN, Scott A.; GUPTA, Sunil; KAMAKURA, Wagner; LU, Junxiang; MASON, Charlotte H. Defection detection: **measuring and understanding the predictive accuracy of customer churn models**. Journal of Marketing Research, v. 43, n. 2, p. 204-211, 2006.

NESPOLO, Daniele, 1986- **Relação entre valor percebido, reputação, confiança e custos de troca como determinantes da retenção de clientes no contexto de serviços de telefonia móvel** Dissertação (Mestrado) – Universidade de Caxias do Sul, Programa de Pós-Graduação em Administração, 2014.

O'MALLEY L. & Tynan, C. (2000). **Relationship marketing in consumer markets rhetoric or reality?** *European Journal of Marketing*, 34(7), 797-815.

PEDRO, Henrique Nobre, **Aplicação de modelos de data mining para previsão de churn em telecomunicações**, 2016

PHILLIP E. Pfeifer, **Non-parametric estimation of mean customer lifetime value**, University of Virginia, Charlottesville, 2005

REGRESSÃO MÚLTIPLA: uma digressão sobre seus usos. Autores: Istvan Karoly Kasznar, PhD Professor Titular da FGV e Presidente da IBCI, Bento Mario Lages Gonçalves, MSc Consultor Senior da IBCI

REICHHELD, F. **O valor da fidelidade** [julho-agosto 2000] HSM Management, ano 4, número 21.

SCHENEIDER, Pedro Henrique. **Análise preditiva de churn com ênfase em técnicas de Machine Learning**: uma revisão, 2016, Dissertação (mestrado) FGV, Escola de Matemática Aplicada.

SILVA, Lauren Paese Martins da Silva. **O MERCADO DE PRODUTOS DIGITAIS: UM ESTUDO DE CHURN DE MIGRADOS DE ASSINATURA DE JORNAL**, UFRGS, 2017

STROUSE, Karen G. **Marketing Telecommunications Services**: new approaches for a changing environment. Norwood: Artech House, 1999. 365 p.

TELECO. INFORMAÇÕES EM TELECOMUNICAÇÕES. **Seção: Banda larga e VOIP**.

Disponível em: <http://www.teleco.com.br/blvoip.asp>. Acesso em 21/09/2020.

_____. **Seção: Telefonia fixa**. Disponível em: <http://www.teleco.com.br/opfixa.asp>. Acesso em 21/09/2020

TRI TELECOM

Disponível em: <<https://www.tri.com.br/sobre-nos/>>

Acesso em 19/09/2020

TROSTER, Roberto Luis; **Introdução à Economia**. São Paulo: Makron Books, 1999.

VAVRA, T. G. **Marketing de Relacionamento – After Marketing**. 4ª ed. São Paulo: Atlas, 1993.

VASQUES, et Al., **Mineração de Textos para Gestão de clientes em empresas de telecomunicações**. XIII Brazilian Symposium of information Systems, Lavras, 2017.

VERHOEF, P. C. **Understanding the effect of customer relationship development**. *Journal of Marketing*, v. 67, n. 4, p. 30-45, 2003.

VILARINHO, Renato Avilez. **Uso de técnicas de mineração de dados para classificação das ocorrências de casos de dengue nos municípios brasileiros** [manuscrito] / Renato Avilez Vilarinho. - 2017.