

INSTITUTO FEDERAL DE SANTA CATARINA

LAYSSA ALVES PACHECO

**Análise de dados aplicada à tomada de decisão
nos negócios.**

**Estudo de caso: Rotatividade de clientes no
ramo de telefonia.**

São José - SC

Maio/2021

**ANÁLISE DE DADOS APLICADA À TOMADA DE
DECISÃO NOS NEGÓCIOS.
ESTUDO DE CASO: ROTATIVIDADE DE CLIENTES
NO RAMO DE TELEFONIA.**

Trabalho de conclusão de curso apresentado à Coordenadoria do Curso de Engenharia de Telecomunicações do campus São José do Instituto Federal de Santa Catarina para a obtenção do diploma de Engenharia de Telecomunicações.

Orientador: Mario de Noronha Neto

São José - SC

Maio/2021

Layssa Alves Pacheco

Análise de dados aplicada à tomada de decisão nos negócios.

Estudo de caso: Rotatividade de clientes no ramo de telefonia./ Layssa Alves Pacheco.
– São José - SC, Maio/2021-

45 p. : il. (algumas color.) ; 30 cm.

Orientador: Mario de Noronha Neto

Monografia (Graduação) – Instituto Federal de Santa Catarina – IFSC

Campus São José

Engenharia de Telecomunicações, Maio/2021.

I. Orientador. II. Coorientador. III. Local. IV. Título

LAYSSA ALVES PACHECO

**ANÁLISE DE DADOS APLICADA À TOMADA DE
DECISÃO NOS NEGÓCIOS.
ESTUDO DE CASO: ROTATIVIDADE DE CLIENTES
NO RAMO DE TELEFONIA.**

Este trabalho foi julgado adequado para obtenção do título de Engenheiro de Telecomunicações, pelo Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina, e aprovado na sua forma final pela comissão avaliadora abaixo indicada.

São José - SC, 20 de Maio de 2021:

Mario de Noronha Neto, Dr.
Orientador
Instituto Federal de Santa Catarina

Ramon Mayor Martins, Ms.
Instituto Federal de Santa Catarina

**Maria Claudia de Almeida Castro,
Dra.**
Instituto Federal de Santa Catarina

RESUMO

A Quarta Revolução Industrial, a qual é baseada na transformação digital, está alterando a dinâmica da sociedade de forma muito rápida, abrupta e disruptiva. Com a chegada do COVID-19, esse processo se acentuou e tem exigido velocidade das pessoas e instituições no progresso de adaptação com a tecnologia. Nesse contexto, o presente projeto tem como objetivo principal abordar a importância da análise de dados para a tomada de decisões mais assertivas e rápidas conforme o cenário exige. Para isso, o projeto tem como objetivo secundário destacar o estudo de caso em uma empresa de telefonia fictícia com relação a rotatividade de clientes, de maneira a exemplificar como a análise de dados pode ser usada para a tomada de decisão nos negócios.

Palavras-chave: Rotatividade de clientes. Transformação digital. Análise de dados.

ABSTRACT

The Fourth Industrial Revolution, which is based on digital transformation, is changing the dynamics of the society in a surprisingly fast, abrupt, and disruptive manner. With the beginning of COVID-19, people and companies demanded urgency in adapting to digital transformation and technology. In this context, this work aims to address the impacts of digital transformation and the benefits of technology, focusing on data analysis as the growth of decision-making in businesses and companies. Consequently, this work aims to highlight the case study in a fictional telephone company regarding customer turnover for the sake of demonstrating how data analysis can support and benefit companies from making decisions.

Keywords: Customer churn. Digital transformation. Data analysis.

LISTA DE ILUSTRAÇÕES

Figura 1 – Tipos de <i>churn</i>	17
Figura 2 – Impacto de diferentes Taxas de Retenção nos resultados financeiros de uma companhia. Fonte: Adaptado de Dirk e LariviÈre (2004).	18
Figura 3 – Gênero vs <i>Churn</i>	30
Figura 4 – Senioridade vs <i>Churn</i>	30
Figura 5 – Matrimônio vs <i>Churn</i>	31
Figura 6 – Dependentes vs <i>Churn</i>	31
Figura 7 – Múltiplas linhas vs <i>Churn</i>	32
Figura 8 – Internet vs <i>Churn</i>	33
Figura 9 – <i>Streaming</i> vs <i>Churn</i>	34
Figura 10 – Tipo de contrato vs <i>Churn</i>	35
Figura 11 – Tempo de cliente vs <i>Churn</i>	36
Figura 12 – Forma de pagamento vs <i>Churn</i>	36
Figura 13 – Índice de satisfação vs <i>Churn</i>	37

LISTA DE TABELAS

Tabela 1 – Gênero.	29
Tabela 2 – Senioridade.	30
Tabela 3 – Matrimônio.	30
Tabela 4 – Dependentes.	31
Tabela 5 – Múltiplas linhas.	32
Tabela 6 – <i>Internet</i>	33
Tabela 7 – <i>Streaming</i>	33
Tabela 8 – Tipo de contrato.	35
Tabela 9 – Tempo de cliente.	35
Tabela 10 – Forma de pagamento.	36
Tabela 11 – Índice de satisfação.	37

LISTA DE ABREVIATURAS E SIGLAS

IDC <i>International Data Corporation</i>	14
CRM <i>Customer Relationship Management</i>	19
IBM <i>International Business Machines Corporation</i>	25

SUMÁRIO

1	INTRODUÇÃO	10
1.1	Motivação	11
1.2	Objetivos	11
1.3	Organização do texto	12
2	FUNDAMENTAÇÃO TEÓRICA	13
2.1	Transformação digital	13
2.2	Rotatividade de clientes em empresas de telefonia	15
2.2.1	<i>Churn</i> : Rotatividade de clientes	15
2.2.2	Relação entre empresas de telefonia e a rotatividade de clientes	19
2.3	Análise de dados aplicada à tomada de decisão nos negócios	20
3	DESENVOLVIMENTO	24
3.1	Metodologia	24
3.2	Métodos Aplicados	27
3.2.1	Análise descritiva	29
3.2.1.1	Dados de perfil	29
3.2.1.2	Dados relacionados a outros serviços contratado	31
3.2.1.3	Dados relacionados ao pagamento	33
3.2.1.4	Dados de satisfação	37
3.2.1.5	Dados selecionados	37
3.2.2	Análise preditiva	38
3.3	Análise dos resultados finais	40
4	CONCLUSÃO	42
	REFERÊNCIAS	43

1 INTRODUÇÃO

Atualmente a sociedade passa pela Quarta Revolução Industrial, a qual é caracterizada por mudanças sociais e industriais devido ao acelerado desenvolvimento tecnológico. Diferente das outras revoluções industriais, a revolução atual chama atenção pela velocidade com que transforma a sociedade. Para identificar isso, basta lembrar como a sociedade se comportava e usufruía de tecnologias há 30 anos em relação aos dias atuais (ARAÚJO et al., 2020) (CARVALHO; TIOSSO; REIS, 2020) (SACOMANO et al., 2018).

A popularização do acesso a tecnologias e a *internet* possibilitaram que a sociedade se mantenha cada vez mais conectada e isso tem mudado a forma com que nos relacionamos e adquirimos conhecimento. A distância física não é mais um impeditivo para o compartilhamento de informações, as comunicações em geral e as conexões entre as pessoas (VILAÇA; ARAÚJO, 2016).

Desde 2020, este cenário tem sido acelerado por causa das necessidades de isolamento pela pandemia do COVID-19. De um dia para o outro, as pessoas precisaram se adaptar à comunicação virtual como principal fonte de informação e relacionamento. Isso transformou os indivíduos, instituições e organizações, pois se antes deste ocorrido, ter estratégias e comportamentos adaptados à era digital era um benefício, com esta mudança radical isso se tornou uma necessidade de sobrevivência (JÚNIOR et al., 2020).

Dessa forma, a Quarta Revolução Industrial firmou seus impactos no mundo. Isso porque o mercado precisou, e ainda precisa, se adaptar muito rápido. A Quarta Revolução exige do ambiente corporativo e industrial que os processos sejam cada vez mais automatizados, devido ao fato que o fluxo de dados é muito grande e o processamento das informações precisam ser consideravelmente mais velozes (SACOMANO et al., 2018).

A velocidade do processo de transformar dados em informações úteis para tomar decisão é fundamental para a sobrevivência dos negócios no mercado. Isso porque atualmente, com o poder da globalização, a competitividade é cada vez mais acirrada e ter a tecnologia ao favor do negócio se tornou uma necessidade e não mais um diferencial (SACOMANO et al., 2018).

Atualmente, 37% da população mundial se encontra no meio digital, sendo que este se tornou o principal formato de comunicação entre as pessoas e as instituições, principalmente com a pandemia (SACOMANO et al., 2018). Estima-se que 3,5% das empresas que não conseguiram se adaptar à pandemia encerraram suas atividades definitivamente, enquanto 31% precisam mudar o formato de trabalho (JÚNIOR et al., 2020). Já quando mudamos o ponto de vista para os benefícios da transformação digital, as empresas consideradas maduras digitalmente são 5 vezes mais lucrativas do que as que não são

(MARTINS et al., 2019).

Com a conectividade constante, as pessoas deixam os seus dados e informações sobre interesses e comportamentos espalhados pela *internet*. Isso acontece em uma rede social, uma busca em sites de pesquisa, ao ouvir uma música, marcar ou realizar uma consulta virtualmente. Para isso, é importante investir em tecnologia e profissionais especializados. Saber extrair, armazenar, manipular, disponibilizar, analisar resultados, ter conclusões e tomar decisões são processos complexos e que requerem conhecimentos diferentes. Por isso, é necessário uma estrutura empresarial baseada em dados que seja bem construída, estruturada e fundamentada (SACOMANO et al., 2018).

Portanto, o presente projeto tem por objetivo abordar a importância da adaptação à transformação digital, destacando como isso pode ser feito e detalhando com a análise de dados é necessária para uma tomada de decisão mais assertiva e rápida.

1.1 Motivação

A principal motivação para o projeto é a atualidade e seriedade da temática da transformação digital no mundo, principalmente diante do contexto da pandemia do COVID-19. Além disso, outro motivador importante é a relevância do tema para profissionais da tecnologia. O mercado, mais do que nunca, está em busca de profissionais técnicos com conhecimento em tecnologias que estejam atualizadas ao mercado, as mudanças sociais e a transformação digital.

1.2 Objetivos

O objetivo principal do projeto é abordar a transformação digital com foco na importância da análise de dados para a tomada de decisão nos negócios. Para desenvolver essa narrativa, pretende-se trabalhar com os seguintes objetivos secundários:

- Descrever como a transformação digital deve ser feita, assim como a implementação do processo de análise de dados.
- Exemplificar o processo de análise através do estudo de caso da rotatividade de clientes de uma empresa fictícia de telefonia. Para isso, teremos os seguintes objetivos específicos:
 - Prever a porcentagem de clientes com tendência ao *churn*.
 - Identificar quais são os principais comportamentos dos clientes que estão relacionados ao *churn*.

1.3 Organização do texto

O texto está organizado da seguinte forma:

- O Capítulo 2 consiste no estudo teórico a respeito dos principais tópicos abordados no projeto, os quais são: uma seção sobre a transformação digital e os seus impactos no mundo (seção 2.1); uma seção para adentrar no contexto do estudo de caso da rotatividade de clientes nas empresas de telefonia (seção 2.2), no qual há uma subseção focada em explicar o que é a rotatividade de clientes (subseção 2.2.1) e outra é voltada a esclarecer a relação entre as empresas de telefonia e a rotatividade de clientes (subseção 2.2.2); por último, uma seção sobre a análise de dados aplicada à tomada de decisão nos negócios (seção 2.3).
- O Capítulo 3 aborda o desenvolvimento da análise de dados de modo a estar dividido em três seções, as quais são: a seção de metodologia, na qual é descrito a metodologia e as etapas iniciais do desenvolvimento (seção 3.1); a seção dos métodos aplicados, que tem por objetivo descrever detalhadamente a análise expondo os resultados coletados (seção 3.2); e, por último, a seção de análise dos resultados cujo propósito é discutir os resultados obtidos de forma críticas com relação à hipótese inicial (seção 3.3).
- O Capítulo 4 traz as considerações finais referente ao projeto.

2 FUNDAMENTAÇÃO TEÓRICA

O presente capítulo visa abordar teoricamente as principais temáticas do projeto, que são a Transformação Digital (seção 2.1), a rotatividade de clientes nas empresas de telefonia (seção 2.2) e a análise de dados aplicada à tomada de decisão nos negócios (seção 2.3). O capítulo tem como objetivo fundamentar o desenvolvimento do projeto a partir do ponto de vista da literatura e dos resultados de outros pesquisadores.

2.1 Transformação digital

Desde o final do século XX, a sociedade vem se transformando juntamente com o aparecimento de novas tecnologias, e este fenômeno é caracterizado como a Quarta Revolução Industrial, também denominada como Indústria 4.0. Diferente das outras revoluções industriais que aconteceram na história, a Indústria 4.0 sucede de maneira muito peculiar devido à sua forma acelerada, abrupta e disruptiva (ARAÚJO et al., 2020) (CARVALHO; TIOSSO; REIS, 2020) (SACOMANO et al., 2018).

O fácil acesso aos dispositivos eletrônicos, como celulares e *tablets*, causado pela redução dos custos dos produtos e matérias-primas diante do desenvolvimento das tecnologias, são uma das razões desta transformação da sociedade. Isso promove e populariza o alcance ao conhecimento e à informação, assim como retroalimenta a cadeia acelerando cada vez mais o desenvolvimento tecnológico (VILAÇA; ARAÚJO, 2016).

O uso dos dispositivos eletrônicos deixaram de ser restritos ao meio acadêmico e/ou de produção para se transformarem em uma nova forma de socialização, alterando a maneira com que nos comunicamos e interagimos (VILAÇA; ARAÚJO, 2016). Esta transformação ocorre por consequência da onipresença digital. Hoje, 37% da população mundial encontra-se conectada digitalmente, e isso tudo é permitido não só pela otimização dos dispositivos eletrônicos, mas também pelo amplo acesso à conexão à *internet* através de avanços como *wi-fi* e redes móveis (SACOMANO et al., 2018).

Por consequência, os principais veículos de informação também estão mudando. Há alguns anos, os mais influentes meios de notícias e disseminação de conhecimento eram canais com jornais, revistas e programas de televisão. A informação era digerida por especialistas e profissionais de forma que os responsáveis pela curadoria escolhiam o que seria popularizado. Contudo, neste novo cenário, a mídia, além de ser majoritariamente digital, também se tornou colaborativa e fluida, de modo que todo indivíduo é passível de ser uma fonte de informação disseminável (SACOMANO et al., 2018) (VILAÇA; ARAÚJO, 2016).

Outra perspectiva e impacto da transformação digital, que também é consequência da migração da comunicação e produção para este meio, é o gerenciamento do grande volume de dados gerados. Em 2015, um estudo da *International Data Corporation* (IDC) afirmou que, 90% dos dados disponíveis na *internet* foram criados nos últimos dois anos (2013-2014) e a velocidade de crescimento prevista consistia em duplicar este volume a cada dois anos (REINSEL; RYDNING; GANTZ, 2020).

Este ponto de vista não considerava o cenário mundial atual, no qual estamos enfrentando a pandemia do COVID-19. Com a necessidade de isolamento, a sociedade precisou acelerar o movimento de adaptação e incorporação das tecnologias no cotidiano. Uma vez que antes, a comunicação presencial ainda era um processo disponível, principalmente para os analfabetos digitais, e com a pandemia esse meio de comunicação se tornou escasso, e em algumas situações inviável. Dessa forma, cada indivíduo e instituição precisou se adequar de acordo com o seu contexto.

De acordo o estudo da IDC lançado no início da pandemia, a expectativa era que mais de 59 ZB de dados fossem criados, copiados e consumidos em 2020. Dessa maneira, só no ano passado seriam criados mais dados do que nos últimos 30 anos, sendo que nos próximos cinco anos esse valor seria três vezes maior devido a velocidade da transformação digital, a resposta e adaptação social à pandemia (REINSEL; RYDNING; GANTZ, 2020).

Portanto, diante desse volume de dados e da mudança de comportamento social, abre-se as portas para que as empresas e negócios possam transformar a forma com que se relacionam com os atuais e futuros clientes. Neste contexto, o relacionamento com os consumidores e a compreensão dos seus comportamentos através dos dados gerados deixou de ser um diferencial competitivo ou uma inovação de mercado para ser o movimento básico da linha de base.

Hoje, principalmente em meio a pandemia, ter a empresa estruturada operacionalmente através de sistemas e processos fundamentados em tecnologia e usufruir disso para tomar decisões baseadas em dados é uma necessidade básica para a sobrevivência no novo mercado, assim como também é uma oportunidade de lucratividade. Segundo o estudo da McKinsey & Company no Brasil, as empresas classificadas como digitalmente maduras são 3 vezes mais lucrativas do que as que não são. Enquanto que, do ponto de vista global, a lucratividade chega a ser 5 vezes maior (MARTINS et al., 2019).

Porém, apesar da urgência com que a situação se apresenta e os benefícios da mudança, 42% das empresas brasileiras não acreditam na importância da tecnologia para a competitividade no mercado e 52% não se beneficia de tecnologias como um diferencial (ARAÚJO et al., 2020). Por isso os impactos da pandemia foram significativos no país, de forma que 31% das empresas mudaram o formato de funcionamento, enquanto que 3,5% encerraram definitivamente as atividades (JÚNIOR et al., 2020). Isso mostra que, a mudança do comportamento empresarial brasileiro se faz necessária de forma ágil,

principalmente para a sobrevivência em um mercado cada vez mais globalizado (SOUZA, 2021).

Para trabalhar o processo de adequação ao meio digital, além de incorporar no ambiente empresarial a ideia de digitalização, e aprimorar e ajustar a comunicação, é preciso investir em tecnologias e profissionais especializados. A estratégia e cultura empresarial deve estar alinhada e fundamentada na tecnologia para se beneficiar e vivenciá-la no dia-a-dia (SACOMANO et al., 2018).

As empresas precisam investir na educação dos profissionais, sejam estes técnicos ou não, assim como os profissionais precisam se adaptar às exigências do mercado de trabalho. Dado que, aprender a se comunicar com base e através dos dados é um dos fundamentos da transformação digital (SACOMANO et al., 2018). Contudo, as estruturas organizacionais também precisam ter os processos e sistemas construídos, estruturados e mantidos de modo a proporcionar que os dados sejam trabalhados funcionalmente. Isso significa garantir os fluxos de gerenciamento de dados, os quais consistem nas fases de coleta, transformação e análise (SACOMANO et al., 2018).

Todas estas partes contribuem em conjunto para a transformação digital das empresas. Contudo, o entendimento do porquê e da importância são os tópicos principais, pois sem esta clareza as demais mudanças não acontecem. A partir do momento que isso é incorporado, é possível coletar dados dos consumidores para entender questões como: seus índices de satisfação, como usufruem os produtos, se são detratores ou promotores das empresas em comunidades virtuais, entre outros pontos de análises possíveis. Entender o comportamento dos clientes e desfrutar disso, legalmente e estrategicamente, permite priorizar e prever assertivamente ações de investimentos em tecnologia, inovação, competitividade e lucratividade.

2.2 Rotatividade de clientes em empresas de telefonia

A análise da rotatividade de clientes em empresas de telefonia é um tópico importante deste projeto, pois é o estudo de caso em questão. Com o intuito de fundamentar o que significa a rotatividade de clientes, qual a importância e os impactos deste tema, principalmente do ponto de vista de empresas de telefonia esta seção visa aprofundar no entendimento teórico do tema (subseção 2.2.1), assim como na relação com o contexto empresarial (subseção 2.2.2).

2.2.1 *Churn*: Rotatividade de clientes

O termo *churn* surgiu da junção das palavras da expressão inglesa "*Change and Turn*", que em tradução livre significa mudar e virar (ALMEIDA, 2010). A expressão, que

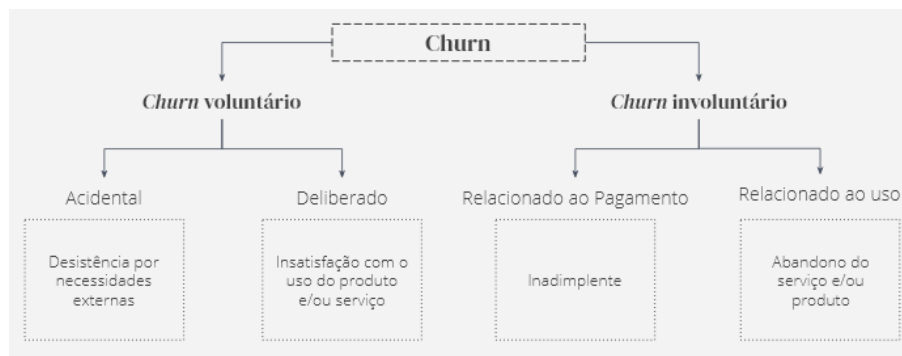
em português é usada como rotatividade de clientes, é utilizada como referência ao processo em que uma empresa perde um cliente para uma empresa concorrente (FERREIRA, 2015).

Em questões numéricas, o *churn* é metrificado a partir da base total de clientes de uma empresa. De forma mais específica, a taxa de *churn* corresponde a porcentagem de clientes que fizeram o distrato no montante de clientes totais em um período de tempo específico. Além disso, o *churn* é classificado em dois tipos diferentes (LU et al., 2012) (FERREIRA, 2015): *churn* voluntário e *churn* involuntário, sendo que o primeiro é subdividido nas categorias de *churn* acidental e *churn* deliberado (ALMEIDA, 2010)(FERREIRA, 2015).

O *churn* involuntário é quando o distrato é feito por parte da empresa. Isso pode acontecer, como explica Ferreira (2015), por inadimplência ou por falta de uso do serviço, sendo que o último é caracterizado pela operadoras de telefonia móvel por clientes em formato pré-pago, os quais não realizam recargas nos últimos três meses. O *churn* voluntário é quando o distrato é feito por solicitação do cliente e pode ocorrer de duas maneiras:

- *Churn* acidental: quando o cliente, por motivos externos, precisa abandonar os serviços que usufrui. Em geral, esse tipo de *churn* não é da vontade do cliente e acontece por motivos como desemprego ou mudança de região (FERREIRA, 2015). A porcentagem do *churn* acidental é pequena, e insignificante, em relação a taxa de *churn* total de uma companhia (ALMEIDA, 2010).
- *Churn* deliberado: quando o cliente decide mudar a prestação de serviço para o concorrente. Esse caso, em específico, se refere a relação cliente e prestador de serviço. Neste contexto, os fatores para impedir o *churn* são associados às definições da empresa sobre o relacionamento com o cliente. De forma que este é o tipo de *churn* que as empresas mais se dedicam a combater (ALMEIDA, 2010). Dentro dessa classificação, Ferreira (2015) menciona dois subtipos de *churn*, os quais são:
 - *Churn* de canal: quando canais de distribuição e comercialização, os quais atendem os contratos dos clientes finais, decidem redirecionar o cliente de uma empresa para o concorrente ao final de um contrato, isso para que possam garantir uma nova comissão.
 - *Churn* promocional: quando um cliente deseja usufruir das vantagens de ser um novo cliente e migra para a empresa concorrente ao final de um contrato.

A Figura 1 torna mais visível o entendimento da classificação do *churn*.

Figura 1 – Tipos de *churn*

Além disso, os principais motivos de *churn* citados por Ferreira (2015) são: a busca dos clientes por tecnologias mais avançadas, acesso a preços melhores e qualidade de serviço superior. Para mais, Ferreira (2015) menciona a influência de fatores psicológicos e sociais, como a interferência de familiares e/ou amigos e associação a marca da empresa concorrente, principalmente em relação a cultura e/ou valores.

A avaliação do *churn*, tanto na saúde de um negócio como em questões financeiras e operacionais, se torna mais preocupante quanto mais maduro é o mercado. Isso porque, como descreve Almeida (2010) na sua pesquisa, quanto maior a maturidade, fluidez e saturação do mercado, mais os consumidores são empoderados para escolher e, conseqüentemente, mais susceptível a mudança os usuários estão, ameaçando assim o lucro das empresas (FERREIRA, 2015).

Diante desse cenário, Almeida (2010) cita o fato de que as estratégias de retenção de clientes e, conseqüentemente, ações para impedir o *churn* são as novas tendências de sobrevivência das companhias no mercado. Por isso é tão relevante o assunto e o investimento por parte das empresa, seja nos departamentos de tecnologia e inteligência de negócio quanto nos projetos de *marketing*, visto que o mercado tende a ser cada vez mais centrado no cliente ao invés de no produto.

Para simplificar a compreensão do impacto do *churn* na receita e sobrevivência de um negócio, Almeida (2010) apresenta um exemplo simples. O intuito é analisar a receita dos próximos 25 anos de uma empresa fictícia no qual cada cliente tem um contribuição líquida de 50 euros por ano. A receita é avaliada comparando três diferentes taxas anuais de *churn*, sendo tais 6%, 7% e 25%. O resultado pode ser visto na Figura 2.

A empresa com *churn* de 6% tem acúmulo de mais de 400 milhões em 25 anos, enquanto que a de 7% tem valor abaixo disso, mas mais próximo do que a empresa com 25% de *churn*, a qual não chega a 200 milhões. Essa simulação avalia apenas o *churn* como influência a fim de destacar o impacto do *churn* a longo prazo no faturamento de uma empresa.

Diante desse cenário, Almeida (2010) apresenta duas categorias para as táticas de

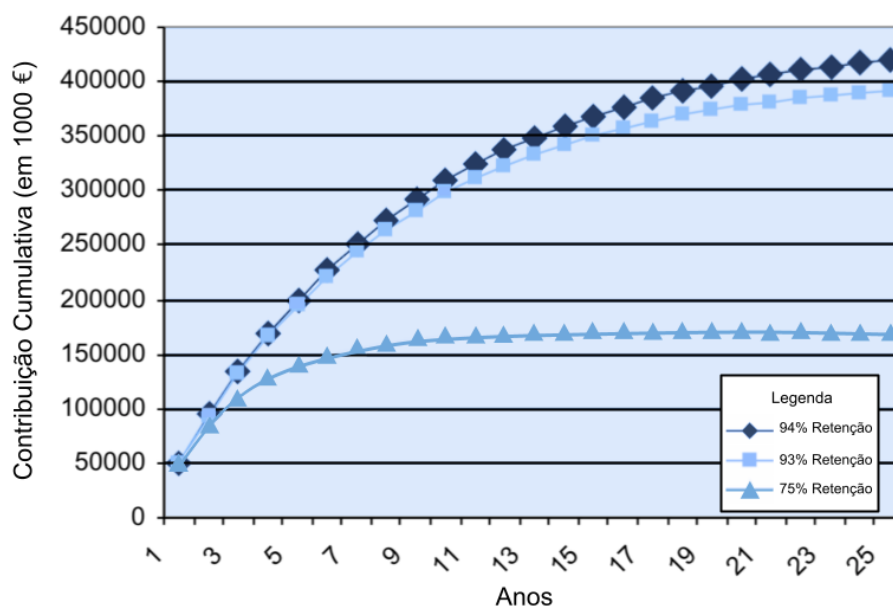


Figura 2 – Impacto de diferentes Taxas de Retenção nos resultados financeiros de uma companhia. Fonte: Adaptado de Dirk e LariviÈre (2004).

gestão do *churn*: a gestão não dirigida e a gestão dirigida, sendo a última subdivida em dois tipos. A gestão não dirigida do *churn* consiste em publicidade em massa, ou seja, sem segmentação de clientes e com o intuito de fortalecer a marca da companhia. Ações como essas não são efetivas com relação ao *churn*, visto que não são focadas nas necessidades dos clientes tendenciosos ao cancelamento de contrato.

Enquanto que a gestão dirigida do *churn* baseia-se na publicidade segmentada para clientes classificados como tendenciosos ao *churn*. Nessa perspectiva, existem dois tipos: estratégias reativas e estratégias pró-ativas. As estratégias reativas são aquelas em que o cliente comunica o desejo de *churn* à empresa e, em contrapartida, recebe algum incentivo para não abandonar os serviços da empresa, o qual, em geral, é um desconto de assinatura. Já as estratégias pró-ativas, são aquelas em que a empresa detecta antecipadamente os clientes de alto risco ao *churn* e de antemão promovem ações de retenção.

Com isso, Almeida (2010) afirma que o aumento do valor econômico devido a retenção de clientes pelas empresas está relacionado aos seguintes fatores:

- Redução da necessidade de aquisição de novos clientes e, portanto, diminuição do valor investido no setor (ALMEIDA, 2010).
- Aumento na quantidade de boas referências devido a satisfação dos clientes e, com isso, construção de uma forma mais barata de aquisição de clientes através das referências (ALMEIDA, 2010).
- Aumento da idade dos clientes na base de consumidores totais e, por consequência,

aumento da receita. Isso com base na comprovação de que clientes mais antigos, tendem a investir mais nos serviços que já utilizam (ALMEIDA, 2010).

2.2.2 Relação entre empresas de telefonia e a rotatividade de clientes

Ao longo da pesquisa bibliográfica é consenso entre os pesquisadores a importância da previsibilidade da rotatividade de clientes nas empresas, principalmente no ramo de telefonia. De acordo com Dalvi et al. (2016), nos últimos anos a indústria de telecomunicações cresceu exponencialmente devido o aprimoramento das tecnologias móveis e as mudanças sociais. Dessa forma, hoje é raro encontrar um pessoa que não tem alguma assinatura com operadora, assim como também não é difícil encontrar clientes que trocam de operadora por diferentes motivos. Cada saída de um consumidor se reflete em custos financeiros, operacionais e de *marketing*.

Atualmente, segundo Mishra e Reddy (2017), o principal objetivo das empresas é reter os atuais clientes devido o custo e dificuldade de adquirir novos. Por isso é relevante entender os clientes e classificá-los entre possíveis *churn* e não *churn*. *Churn* é o termo em inglês para rotatividade de clientes. No caso em específico, seriam os clientes que deixaram de ser consumidores da empresa (*churn*) e os que continuam como consumidores (não *churn*).

Lu et al. (2012) afirma que a indústria de telecomunicações tem uma taxa média mensal de *churn* de 2,2%, ou seja, a cada 50 assinantes de um serviço, ao menos um troca de operadora durante o mês. Se observarmos da perspectiva de Dalvi et al. (2016), o qual afirma que o ramo de telefonia móvel lida com uma quantidade de cliente em torno de bilhões de usuários, o distrato mensal de 2,2% deles representa uma receita financeira significativa que precisa ser prevista e administrada.

Quando o assunto é relacionado as operadoras de telefonia móvel, Mishra e Reddy (2017) afirma que os celulares tem um papel significativo na sociedade atual. Portanto, da mudança do mercado e da demanda mundial pelo setor, causadas pelas saturação, competitividade acirrada e a portabilidade, as empresas precisaram e continuando precisando se reinventar, como complementa Dalvi et al. (2016). Isso promove o uso de artifícios como a variedade de ofertas de assinaturas com base em diferentes valores e serviços por parte das empresas, assim como a adoção do *Customer Relationship Management* (CRM).

Dentre as pesquisas bibliográficas, Dalvi et al. (2016) é o autor que mais debate sobre a importância da predição do *churn* em relação ao seus efeitos no mercado. Além disso, o autor ressalta a necessidade de um modelo capaz de fomentar os motivos de *churn* ao invés de apenas agrupar clientes entre os estado de possíveis *churn* e não *churn*. Isso porque, na visão de Dalvi et al. (2016), o *marketing* do mercado precisa se munir de dados para projetar políticas e táticas para estratégias de retenção. Apenas dessa maneira, de

acordo com Dalvi et al. (2016), é possível reduzir custos e danos.

Sem uma lista de clientes identificados de acordo com os principais motivos que os levariam ao *churn*, para Dalvi et al. (2016), as ações de *marketing* seriam individuais e aleatórias, sem condizer com as estratégias da empresa. Pois, trabalhos individuais são custosos, demorados e não oferecem visibilidade do retorno de investimento. Enquanto que, ações de *marketing* estratégicas com grupos definidos e caracterizados são mais baratas e efetivas. Dessa maneira, aumenta-se a receita de permanência dos cliente e redução de gastos, principalmente no ponto de vista de aquisição de novos clientes, como também concorda Mishra e Reddy (2017).

Complementando a visão de Dalvi et al. (2016), Lopes (2007) aponta estudos que correlacionam positivamente a relação entre índices de retenção e lucratividade, valor do cliente para a empresa em questões de lealdade e satisfação e aumento do valor da própria empresa. Ou seja, investigar e promover ações de retenção dos clientes é comprovadamente associado a diferentes fontes de lucratividade.

Para gerenciar os clientes existem plataformas de CRM, as quais já são usadas por diferentes setores de prestação de serviço. Dalvi et al. (2016) cita como exemplo de uso as operadoras de cartão de crédito que além de gerir os clientes também usam as informações do CRM para entender e prever comportamentos de inadimplência e *churn*. O CRM consegue produzir um grande banco de dados com ricas informações a respeito dos clientes e sua satisfação. Dessa forma, o CRM possibilita análises em diferentes níveis, como a previsibilidade de *churn* e segmentação de clientes. Além de facilitar o gerenciamento dos consumidores por si só, como o próprio nome da ferramenta já deixa claro. Além disso, A e Nestor (2018) apresentam, além dos pontos anteriores, a perspectiva de que o CRM pode ser usado para identificar oportunidades de melhoria da qualidade de serviço, do atendimento e da experiência do cliente.

A e Nestor (2018) alega que a retenção de clientes está intimamente relacionada com a qualidade da experiência e que, portanto, este é o principal influenciador da receita impactada pelo *churn*. A qualidade da experiência é medida de acordo com a nota atribuída pelo cliente com base na satisfação do serviço prestado. Com isso, A e Nestor (2018) defende o uso de análise de dados, aprendizado de máquina e inteligência artificial como ferramentas para encontrar os padrões de comportamento capazes de influenciar o *churn* e a satisfação dos clientes. A e Nestor (2018) ainda reforça a necessidade e aplica em sua pesquisa o pré-processamento dos dados do CRM antes do uso.

2.3 Análise de dados aplicada à tomada de decisão nos negócios

A análise de dados aplicada à tomada de decisão envolve questões mais profundas do que apenas definir o tipo de análise e os requisitos de variáveis. A incorporação desse

processo e, conseqüentemente, a geração de impacto nos negócios exige o envolvimento de toda a empresa através da cultura empresarial (DÍAZ; ROWSHANKISH; SALEH, 2018).

Segundo estudos da McKinsey & Company, ser uma empresa digitalmente madura impacta na lucratividade e competitividade no mercado e é uma necessidade de adaptação diante da pandemia do COVID-19 (MARTINS et al., 2019) (SNEADER; STERNFELS, 2020). Além disso, os processos se tornam mais ágeis e as tomadas de decisão menos arriscadas. Contudo, para se transformar nesse tipo de empresa não é fácil e precisa partir de raízes que norteiam a cultura empresarial, tornando-a uma cultura guiada por dados (DÍAZ; ROWSHANKISH; SALEH, 2018).

A transição de sair de uma cultura não guiada por dados para uma guiada por dados precisa ser construída e baseada em mudanças faseadas. Essa mudança não pode ser radical, porque isso gera segregação entre pessoas e interesses, o que leva a falsa sensação de fracasso já no início do projeto. A implementação precisa ser norteada em propósitos claros e bem comunicados para toda a empresa (DÍAZ; ROWSHANKISH; SALEH, 2018).

Essa é uma transformação estrutural que envolve alterações de processos e condução de resultados, de forma que não pode ser um projeto isolado de times de engenharia e desenvolvimento de produto, mas de toda a empresa. Isso significa envolver desde os profissionais mais técnicos, até gestores e profissionais das áreas de negócios (DÍAZ; ROWSHANKISH; SALEH, 2018).

A criação dos dados é a primeira etapa do fluxo que leva a tomada de decisão e isso acontece através da rotina de processos dos diferentes profissionais dentro da empresa. Isso se caracteriza pela documentação dos pontos importantes de uma conversa com um cliente, pelos dados fornecidos por alguém interessado nos produtos/serviços da empresa, pelos dados gerados pelos usuários de um sistema, dentre outras situações (SACOMANO et al., 2018).

Cada sistema usado por um profissional ou consumidor é uma fonte de geração de dados. Portanto, os agentes de cada um dos processos precisam entender e absorver a importância de fornecer cada dado dentro do processo geral. Caso contrário, o fornecimento dos dados pode falhar pelo simples fato de que quem os gera não enxerga valor nisso. Por isso, a relevância em deixar claro o porquê desta transformação digital para uma cultura fundamentada em dados (DÍAZ; ROWSHANKISH; SALEH, 2018).

Depois da geração dos dados, as etapas técnicas de coletar, armazenar e transformar os dados é a principal entrega de valor dos profissionais especializados. Isso porque, os dados são transformados em informações úteis e, portanto, passíveis de serem usados para tomada de decisão (SACOMANO et al., 2018). Porém, os profissionais técnicos e de negócios precisam estar alinhados conceitualmente sobre os significados de cada etapa e indicadores, caso contrário, as informações transmitidas podem não ser confiáveis e/ou

refletirem a realidade do processo ou contexto (DÍAZ; ROWSHANKISH; SALEH, 2018).

Outro ponto importante é que uma cultura guiada por dados é uma cultura de tomada de decisão. Consequentemente, esses dois pontos estão intimamente relacionados e não existe separação entre tais. Uma vez que, o objetivo principal da coleta, análise e implementação dos dados é tomar uma decisão. Por isso, é primordial a comunicação entre as áreas e os profissionais (DÍAZ; ROWSHANKISH; SALEH, 2018).

Ter dados disponíveis e não transformá-los em informações úteis torna o processo sem valor. Não é estratégico coletar dados sem objetivos claros do porquê, visto que sem essa definição os dados são apenas números sem importância. Este é um dos problemas de provar o valor da cultura de dados, porque quando a estratégia não é bem definida e comunicada, os dados não conseguem guiar. Com isso, não há resultados e, consequentemente, não se enxerga o valor de continuar a promover a transformação digital (DÍAZ; ROWSHANKISH; SALEH, 2018).

A partir do momento que esse cenário é compreendido e a empresa pode seguir para um próximo passo, adentra-se nos aspectos da análise de dados em si. Nesta questão, a qual assemelha-se a um projeto, tem como estágios de execução delimitar o escopo, definir os requisitos, escolher o tipo de análise apropriada, executar o processo da análise e expor/comunicar os resultados.

Dedicar tempo delimitando o escopo e definindo os requisitos são essenciais para a condução e sucesso das etapas seguintes. Em geral, essas fases exigem muita comunicação e alinhamento de expectativas, de forma que parecem não gerar um valor palpável. Todavia, quando estes passos não são bem trabalhados impactam nos resultados finais, no tempo de execução e na percepção da entrega de valor.

Resumidamente, a delimitação do escopo engloba o entendimento do problema que se deseja resolver de modo a compreender a fundo a causa raiz da dor. Enquanto que, a definição dos requisitos relaciona-se com a elucidação das premissas e cenários que serão considerados, ou não, para o desenvolvimento do projeto de análise. Além disso, essa etapa tem por objetivo alinhar a expectativa de quem solicita a análise com quem executará, para que todos estejam cientes do processo e dos resultados esperados.

Após este período, deve-se escolher a análise de dados adequada para o cenário. Uma vez que, quando abordamos análise de dados aplicada à tomada de decisão nos negócios restringimos este contexto para três tipos específicos, os quais são (XAVIER, 2018) (EVANS; LINDNER, 2012):

- **Análise descritiva:** é o tipo de análise inicialmente aplicada pelas empresas. Este tipo consiste em analisar os dados para identificar comportamentos passados e atuais para tomar decisões a partir destes cenários. Em geral, usa técnicas de estatísticas simples,

análises de clusterização e associação para descrever e monitorar os resultados e indicadores em relatórios e painéis de visualização de dados.

- **Análise preditiva:** é o tipo de análise que usa dados passados e atuais a fim de detectar padrões, relacionamentos e tendências. Diferente na análise descritiva, a análise preditiva é capaz de reconhecer riscos, agrupamentos e padrões em grande volumes de dados com relação a cenários futuros. Isso é feito através de técnicas de regressão e redes neurais para desenvolver previsões e simulações.
- **Análise prescritiva:** é o tipo de análise que sugere alternativas otimizadas para alguma necessidade específica, como por exemplo, o valor ideal para um produto ou o desconto máximo permitido em uma venda sem que haja prejuízo. Um método recorrentemente citado na literatura é o Monte Carlo.

Depois de escolher a análise ideal, é o momento de executar. A análise descritiva permite que seja executada por algoritmos mais simples através da linguagem Python ou R. Assim como também possibilita ser feita por relacionamentos de dados em *softwares* disponíveis no mercado (Power BI, Tableau, IBM Watson Analytics, Data Studio) ou em formato *open source* (BIRT), no qual o foco seja promover a visualização dos dados em formato de painéis e/ou relatórios.

No caso da análise preditiva, isso envolve algoritmos mais complexos e, em geral, não se utiliza ferramentas pagas. Isso porque, são áreas de desenvolvimento da inteligência artificial muito exploradas por acadêmicos, profissionais e entusiastas que promovem e investem em modelos *open source*. Alguns exemplos para análises preditivas podemos citar a Regressão Logística e a Árvore de Decisão.

Com a análise definida e executada, é necessário comunicar e explicar os resultados para que a partir deste ponto decisões sejam tomadas. Para isso é preciso uma comunicação alinhada entre as áreas e os profissionais para que todos entendam a linguagem e os termos usados.

Neste trabalho foram usados a análise descritiva com uso do uso do *software* Power BI, e a análise preditiva com a aplicação das técnicas de Regressão Logística e Árvore de Decisão, sendo ambas através da linguagem Python e ambiente do Google Collab. No Capítulo 3 são pincelados conceitos a respeito das análises preditivas, porém como esses são métodos de aprendizado de máquina bem explorados na literatura, optou-se por não aprofundar sobre o tema na fundamentação teórica. De toda forma, informações teóricas adicionais podem ser encontradas em Schneider (2016) e Laboratório de Aprendizado de Máquina em Finanças e Organizações (2017).

3 DESENVOLVIMENTO

Diante do contexto passado, no qual descrevemos a importância da análise de dados para a tomada de decisão, o presente capítulo visa exemplificar como a análise de dados pode ser usada para a tomada de decisão nos negócios. Para isso, analisar-se-á a rotatividade de clientes em uma empresa fictícia de telefonia demonstrando como esse processo contribui para a tomada de decisão. Para orientar o desenvolvimento, o capítulo está dividido em:

- Metodologia (seção 3.1): destina-se a explicar o processo de análise, mas focando nas etapas de delimitação do escopo, definição dos requisitos e escolha do tipo da análise;
- Métodos aplicado (seção 3.2): tem por objetivo detalhar o passo-a-passo de cada tipo de análise executada e apontar seus respectivos resultados. Nesta etapa não serão discutidos os resultados, apenas apresentá-los.
- Análise dos resultados (seção 3.3): tem a finalidade de concluir a respeito dos resultados de ambas as análises, de modo a correlacioná-los. Além disso, pretende-se interpretar estes resultados diante da hipótese inicial.

3.1 Metodologia

O processo de desenvolvimento da análise consistiu nas seguintes etapas:

1. Delimitação do escopo.
2. Definição dos requisitos.
3. Escolha do tipo da análise.
4. Execução da análise.
5. Discussão da análise

A delimitação do escopo refere-se a escolha do tema da análise, que é investigar a rotatividade de clientes em uma empresa de telefonia fictícia. Neste ponto dois tópicos são abordados, a temática da rotatividade de clientes e o contexto da empresa de telefonia. O fato dos dados serem fictícios é uma parte da etapa de definição de requisitos.

O ponto de partida para a delimitação do escopo foi decidir explorar uma dor relacionada à empresas de telefonia, pois se trata de uma indústria com temas muito trabalhos ao longo do curso de Engenharia de Telecomunicações. Com isso em mente, optou-se pelo principal atrito nesse mercado, que é a rotatividade de clientes, a qual foi agravada pela portabilidade, conforme mencionado anteriormente.

O *churn* é uma problemática antiga do mercado de telefonia, que foi acentuado com a chegada da pandemia do COVID-19. Isso porque o uso dos serviços de telefonia se tornaram cruciais nessa situação. Dessa forma, por mais que esse seja um problema do cenário pré-pandemia, ainda assim continua atual. Porém, é importante destacar que as bases envolvidas na análise foram concebidas anteriormente ao COVID-19.

Na etapa de definição dos requisitos, estabeleceu-se que para a análise seriam usados dados fictícios. Uma vez que, o objetivo dessa parte do projeto é demonstrar como a análise de dados apoia e contribui para a tomada de decisão. Portanto, o uso de dados fictícios cumpri o requisito do objetivo do projeto, que é fazer uma análise de rotatividade de clientes em uma empresa de telefonia, assim como evita a complexidade ética de manipular dados sigilosos de usuários reais.

Além disso, nessa fase escolheu-se as bases de dados a serem usadas no desenvolvimento. Optou-se por duas bases disponibilizadas pela empresa *International Business Machines Corporation* (IBM), as quais são denominadas *Telco Customer Churn* e *Telco Customer Churn Status* e podem ser acessadas através do portal IBM (2019). Como os dados são fictícios, é importante reforçar que as bases não possuem dados de usuários reais. Contudo, os dados são capazes de caracterizar clientes consumidores de serviços de uma empresa de telefonia (IBM, 2020).

Os dados presentes na *Telco Customer Churn* são relacionados ao *status* do cliente na empresa e aos serviços utilizados, enquanto que os dados da *Telco Customer Churn Status* apresenta dados adicionais de *status* do cliente. Em conjunto, podemos categorizar os dados da seguinte maneira:

- Dados geográficos;
- Dados de perfil do usuário;
- Dados de serviços usados;
- Dados relacionados à pagamento;
- Dados referentes a rotatividade dos cliente.

As bases de dados são usadas pela IBM para incentivar o estudo e aprendizado por clientes, profissionais e curiosos na área de análise de dados e *machine learning*. O foco

principal é promover o uso dos seus *softwares* e sistemas dentro do mercado (IBM, 2017). As bases disponibilizadas também são amplamente usadas em sites de compartilhamento de base de dados com foco em estudos relacionados à ciência de dados, como Kaggle, por exemplo.

Após a seleção das bases de dados, iniciou-se a terceira etapa do processo de desenvolvimento da análise, a qual corresponde à escolha do tipo da análise. Para definir o tipo é necessário considerar os objetivos específicos da análise, os quais buscam resultados que, em geral, são numérico. Portanto, como o objetivo principal da análise é investigar a rotatividade de clientes em uma empresa de telefonia fictícia, os objetivos secundários são:

1. Prever a porcentagem de clientes com tendência ao *churn*.
2. Identificar quais são os principais comportamentos dos clientes que estão relacionados ao *churn*.

Em vista disso, e de acordo com os tipos de análises apresentados no capítulo anterior do projeto, entende-se que os tipos de análise mais adequados para atingir os resultados dos objetivos específicos são: a análise descritiva e a análise preditiva.

Com a escolha do tipo da análise, seguiu-se para a etapa de execução. Contudo, antes de executar, é preciso definir alguns requisitos técnicos relacionados às análises, assim como o passo-a-passo da execução. Para a análise descritiva, inicialmente usou-se o *software* de *business intelligence* Power BI da empresa Microsoft. O Power BI é uma ferramenta focada na visualização de dados e permite identificar o comportamento e relacionamento entre variáveis de forma rápida e simples, após a estruturação dos dados.

Com a manipulação dos dados no Power BI, selecionar-se-á aqueles que serão usados na análise preditiva. Isso porque, existem dados que podem tendenciar o resultado de previsão de *churn* caso apresentem desproporcionalidade com relação a esse indicador. Essas situações podem ser identificadas na análise descritiva. Além disso, para afirmar quais são os principais comportamentos que estão relacionados ao *churn* far-se-á uma análise de correção entre os dados usando os mesmos meios da análise preditiva.

Com relação a análise preditiva, as técnicas escolhidas e avaliadas no processos foram: Regressão Logística e Árvore de Decisão. De forma que, a triagem das técnicas se baseou nos seguintes critérios:

- Simplicidade de aplicação.
- Familiaridade pessoal com o método.
- Diversidade de conteúdo disponível sobre o assunto.

De maneira resumida a Regressão Logística é uma técnica de aprendizado de máquina supervisionado, na qual a partir de variáveis independentes prevê-se variáveis dependentes após o treinamento do algoritmo. A técnica identifica quais variáveis independentes influenciam no resultado da variável dependente e como a influência acontece. Dessa forma, o modelo mede a relação entre as variáveis estimando a probabilidade entre elas, de modo a restringir os resultados de saída para um intervalo de probabilidade entre 0 e 1. As vantagens desta técnica são: simplicidade de executar, facilidade de interpretação e nível de acurácia satisfatória considerando o grau de complexidade (SCHNEIDER, 2016) (GEVERT, 2009). Para mais informações, conforme mencionado no capítulo anterior, recomenda-se a leitura de Laboratório de Aprendizado de Máquina em Finanças e Organizações (2017).

Enquanto isso, de modo abreviado, a Árvore de Decisão é um dos modelos mais conhecidos de aprendizado de máquina por ser intuitivo e de fácil compreensão e aplicação. Esse método se encaixa nos algoritmos de aprendizado de máquina supervisionado, assim como a Regressão Logística. O intuito do método é dividir sucessivamente o conjunto de dados de entrada que se encontram na "raiz" de forma sucessiva até as "folhas", assemelhando-se a estrutura de uma árvore. Para isso, há os nós internos, os quais classificam os atributos e subdividem o conjunto em subconjuntos. As subdivisões em relação aos atributos continuam até que se atinjam as "folhas", que são as classificações alvo (NETO, 2018). Para mais informações, conforme mencionado no capítulo anterior, recomenda-se a leitura de Laboratório de Aprendizado de Máquina em Finanças e Organizações (2017) e Schneider (2016).

Por fim, com o intuito de conduzir a análise preditiva, escolheu-se desenvolver um código na linguagem Python usando como ferramenta o Google Colab da empresa Google. As demais etapas de execução, tanto da análise descritiva, quanto da preditiva, serão abordados em detalhes, respectivamente, na subseção 3.2.1 e na subseção 3.2.2. Enquanto que, a etapa seguinte da análise, que consiste em discutir os resultados, será elaborada na seção 3.3.

3.2 Métodos Aplicados

Diante das etapas para análise de dados apresentadas anteriormente, a presente seção visa detalhar a fase de execução da análise. Para a condução desse processo, independente do tipo de análise, seguiu-se as seguintes sub-etapas, :

1. Unificação das bases de dados.
2. Limpeza da base unificada.
3. Aplicação dos métodos.

4. Coleta dos resultados.

Com a escolha das bases de dados, conforme detalhado na seção 3.1, fez-se o agrupamento dos dados em uma única base manualmente. Isso foi feito através da chave identificadora dos clientes, que eram iguais entre as bases, demonstrando assim que se tratavam dos mesmos clientes. Nessa etapa, a base foi trabalhada usando planilhas e fórmulas.

Em seguida, realizou-se a limpeza dos dados de forma a remover aqueles considerados repetidos ou que não seriam úteis a pesquisa. Nesse caso, é importante lembrar que a análise tem por objetivo trabalhar com a rotatividade dos clientes, prevendo a tendência de *churn* e os principais comportamentos relacionados à isso (com o objetivo de exemplificar o uso da análise de dados para a tomada de decisão). Dessa forma, para definir o que são dados úteis teve-se ao fato de que a literatura afirma que, o *churn* está mais relacionado à satisfação e ao uso dos serviços e/ou produtos por parte do cliente (A; NESTOR, 2018). Portanto, os dados que não agregavam a análise e que não são relativos aos aspectos citados foram removidos.

Dentre os dados listados como não necessário estão: dados geográficos, dados de pagamento com uso de papel e datas de referência dos dados. Esse processo foi executado de maneira diferente dependendo do tipo de análise. Os dados analisados descritivamente foram removidos ao longo do processo de uso do *software* Power BI. Enquanto que, os dados analisados preditivamente foram removidos através da função *drop* em Python.

Logo após, categorizou-se os dados da seguinte maneira:

- Dados de perfil: gênero, senioridade, matrimônio e dependentes.
- Dados relacionados a outros serviços contratados: uso de múltiplas linhas telefônicas, *internet* e/ou *streaming*.
- Dados relacionados ao pagamento: tipo de contrato, tempo de contrato e modelo de pagamento;
- Dados de satisfação: índice de satisfação do cliente;
- Dados relacionado ao *Churn*: categorização e motivo;
- Métrica de peso: valor numérico definido a partir do cálculo relacionado à importância de um cliente com base em diferentes critérios pré-definidos.

Com a base de dados definida e tratada corretamente, executou-se as etapas seguintes, que são: aplicação dos métodos e coleta dos resultados. Conforme dito na seção 3.1, a análise descritiva usará o *software* Power BI para visualizar e relacionar os

dados de cada categoria com o *churn*. A partir dos resultados coletados, selecionar-se-á aqueles que demonstrarem mais relação com o *churn* para serem usados na análise preditiva. Enquanto que, a análise preditiva, inicialmente avaliará as técnicas de Regressão Logística e Árvore de Decisão em relação a acuracidade. A técnica com melhor índice de acuracidade será aplicada na base de dados tratada para obter a quantidade e porcentagem de clientes com tendência ao *churn*.

3.2.1 Análise descritiva

A análise descritiva trabalhou com quatro agrupamentos, conforme descrito na seção 3.2. Os detalhes da análise estão especificados nas seções seguintes.

3.2.1.1 Dados de perfil

Os dados de perfil trabalhados foram:

- Gênero: classifica a amostra entre indivíduos identificados como homens ou mulheres.
- Senioridade: classifica a amostra entre indivíduos com 65 anos ou mais e indivíduos com menos de 65 anos.
- Matrimônio: classifica a amostra entre indivíduos casados ou não casados.
- Dependentes: classifica a amostra entre indivíduos com dependentes ou sem dependentes.

Com relação ao gênero, identificou-se que a base consiste em uma amostra de 50% homens e 50% mulheres, de acordo com a Tabela 1.

Gênero	N° Clientes	% N° de Clientes
Masculino	3.555	50%
Feminino	3.488	50%
Total	7.043	100%

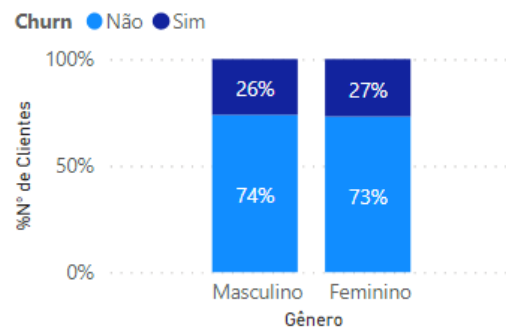
Tabela 1 – Gênero.

Acerca dos homens, 26% deles são classificados como *churn*. Enquanto que, dentre as mulheres 27% são classificadas como *churn*. Isso é na visto Figura 3.

Portanto, não é possível identificar uma relação simples entre o *churn* e o gênero. Por isso, estes dados não serão usados na análise preditiva.

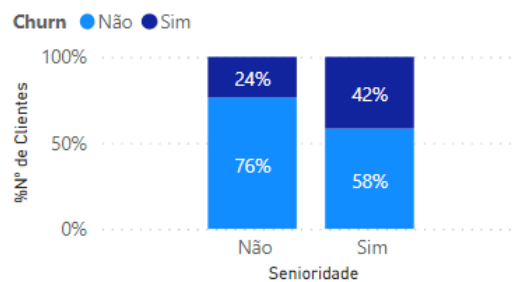
Com relação a senioridade, identificou-se que a base consiste em uma amostra de 16% de clientes sêniores e 84% de clientes não sêniores, conforme Tabela 2.

Quanto aos clientes sêniores, 42% são classificados como *churn*. Enquanto que os clientes não seniores são 24% classificados como *churn*. Isso é visto na Figura 4

Figura 3 – Gênero vs *Churn*.

Senioridade	Nº Clientes	% Nº de Clientes
Não	5.901	84%
Sim	1.142	16%
Total	7.043	100%

Tabela 2 – Senioridade.

Figura 4 – Senioridade vs *Churn*.

Diante destes dados, podemos reconhecer que há uma possível relação entre o cliente sêniores e o *churn*. Contudo, é importante notar que a amostra é significativamente pequena para ser conclusiva. Por isso, estes dados não serão usados na análise preditiva.

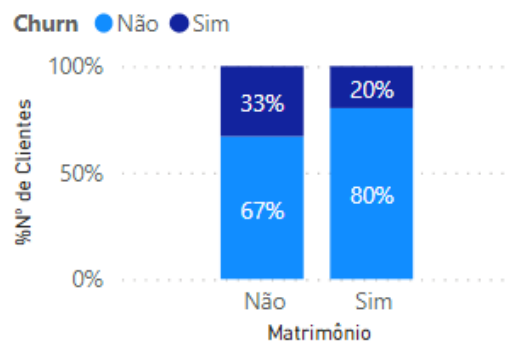
Com relação ao matrimônio, observou-se que a base consiste em uma amostra de 48% de clientes casados e 52% de clientes não casados, conforme Tabela 3.

Matrimônio	Nº Clientes	% Nº de Clientes
Não	3.641	52%
Sim	3.402	48%
Total	7.043	100%

Tabela 3 – Matrimônio.

No tocante aos clientes casados, 20% são classificados como *churn*. Enquanto que os clientes não casados são 33% classificados como *churn*. Isso é visto na Figura 5.

Com isso, podemos apontar que há uma possível relação entre os clientes não casados e o *churn*. Por isso, estes dados serão usados na análise preditiva.

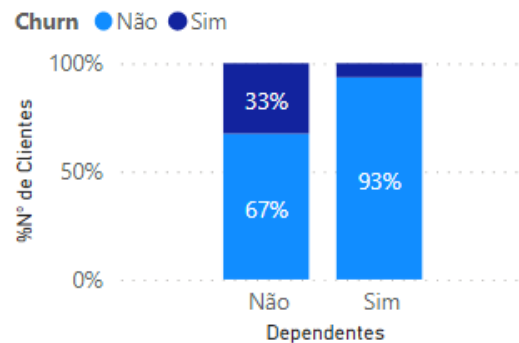
Figura 5 – Matrimônio vs *Churn*.

Com relação aos dependentes, identificou-se que a base consiste em uma amostra de 23% de clientes com dependentes e 77% de clientes sem dependentes, conforme Tabela 4.

Dependentes	Nº Clientes	% Nº de Clientes
Não	5.416	77%
Sim	1.627	23%
Total	7.043	100%

Tabela 4 – Dependentes.

Dos clientes com dependentes, 7% são classificados como *churn*. Enquanto que os clientes sem dependentes são 33% classificados como *churn*. Isso é visível na Figura 6.

Figura 6 – Dependentes vs *Churn*.

Diante destes dados, podemos reconhecer que há uma possível relação entre o cliente sem dependentes e o *churn*. Contudo, é importante notar que a amostra é significativamente tendenciosa para ser conclusiva, visto que a maioria dos clientes são aqueles com maior representatividade de *churn*, o que é de se esperar em uma amostra. Por isso, estes dados não serão usados na análise preditiva.

3.2.1.2 Dados relacionados a outros serviços contratado

Os dados analisados com relação ao uso de outros serviços contratados foram:

- Uso de múltiplas linhas telefônicas: classifica a amostra entre indivíduos que usam mais de uma linha telefônica ou apenas uma linha telefônica.
- *Internet*: classifica a amostra entre indivíduos que contrataram ou não o plano de *Internet*.
- *Streaming*: classifica a amostra entre indivíduos que contrataram ou não os planos *streaming*, que podem ser de TV e/ou filmes.

Com relação ao uso de múltiplas linhas, observou-se que a base consiste em uma amostra de 42% de clientes que são usuários de múltiplas linhas e 58% de clientes que não são usuários de múltiplas linhas, conforme Tabela 5.

Múltiplas Linhas	Nº Clientes	% Nº de Clientes
Não	4.072	58%
Sim	2.971	42%
Total	7.043	100%

Tabela 5 – Múltiplas linhas.

Desta forma, com relação aos clientes que são usuários de múltiplas linhas, 29% são classificados como *churn*. Enquanto que os clientes que não são usuários de múltiplas linhas são 25% classificados como *churn*. Isso é mostrado na Figura 7.

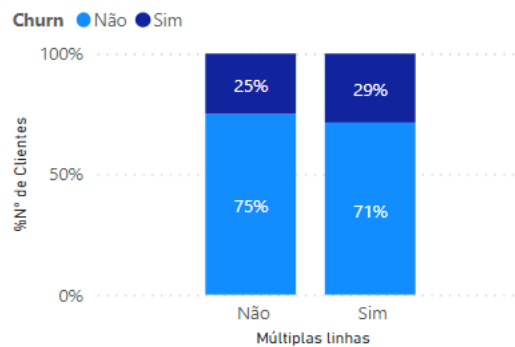


Figura 7 – Múltiplas linhas vs *Churn*.

Portanto, não é possível identificar uma relação simples entre o *churn* e o uso ou não uso de múltiplas linhas. Por isso, estes dados não serão usados na análise preditiva.

Com relação ao uso de plano de *internet*, notou-se que a base consiste em uma amostra de 78% de clientes que são contratantes e 22% de clientes que não são contratantes, conforme Tabela 6.

Segundo os clientes que são contratantes de *internet*, 32% são classificados como *churn*. Enquanto que os clientes que não são contratantes são 7% classificados como *churn*. Isso é mostrado na Figura 8.

Internet	Nº Clientes	% Nº de Clientes
Não	1.526	22%
Sim	5.517	78%
Total	7.043	100%

Tabela 6 – Internet

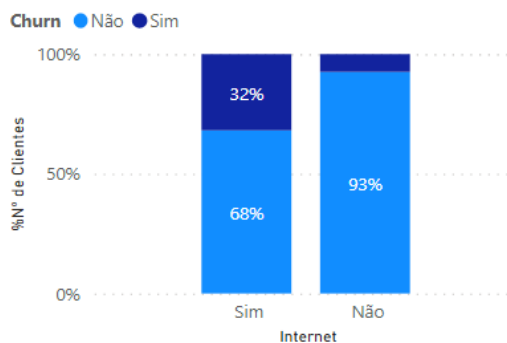


Figura 8 – Internet vs Churn.

Diante destes dados, podemos reconhecer que há uma possível relação entre o cliente que são contratantes de *internet* e o *churn*. Contudo, é importante notar que a amostra é significativamente tendenciosa para ser conclusiva, visto que a maioria dos clientes são aqueles com maior representatividade de *churn*, o que é de se esperar em uma amostra. Por isso, estes dados não serão usados na análise preditiva.

Com relação ao uso de *streaming*, identificou-se que a base consiste em uma amostra de 50% de clientes que são usuários de *streaming* e 50% de clientes que não são usuários de *streaming*, conforme Tabela 7.

Streaming	Nº Clientes	% Nº de Clientes
Não	3.544	50%
Sim	3.499	50%
Total	7.043	100%

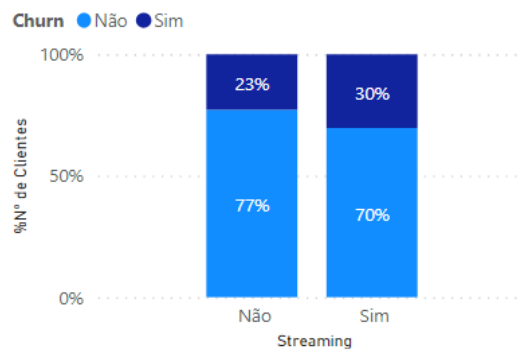
Tabela 7 – Streaming

Quanto aos clientes que são usuários de *streaming*, 30% são classificados como *churn*. Enquanto que os clientes que não são usuários de *streaming* são 23% classificados como *churn*. Isso é demonstrado na Figura 9.

Diante destes dados, podemos identificar que há uma possível relação entre o cliente que são usuários de *streaming* e o *churn*. Por isso, estes dados serão usados na análise preditiva.

3.2.1.3 Dados relacionados ao pagamento

Os dados analisados com relação ao pagamento foram:

Figura 9 – *Streaming vs Churn.*

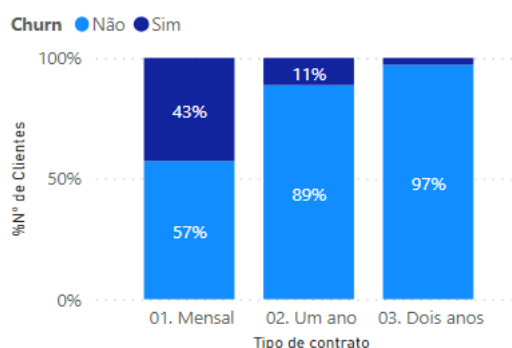
- Tipo de contrato: classifica a amostra entre indivíduos que contrataram o serviço e/ou produto para pagamento mensal, pagamento antecipado de um ano ou pagamento antecipado de dois anos.
- Tempo de contrato: classifica a amostra em relação ao tempo em que os usuários usufruem dos serviços e/ou produtos da empresa. A categoria considera períodos de menos de seis meses até acima de quatro anos.
- Modelo de pagamento: classifica a amostra entre indivíduos que contrataram o serviço e/ou produto para pagamento com cartão de crédito com cobrança automática, pagamento eletrônico, pagamento em cheque ou pagamento em transferência bancária automática.

Com relação ao tipo de contrato, observou-se que a base consiste em uma amostra de 55% de clientes que pagam mensalmente, 24% de clientes que pagam antecipado por um período de dois anos e 21% de clientes que pagam antecipado por um período de um ano, conforme Tabela 8.

Tipo de Contrato	N° Clientes	% N° de Clientes
01. Mensal	3.875	55%
02. Um ano	1.473	21%
03. Dois anos	1.695	24%
Total	7.043	100%

Tabela 8 – Tipo de contrato.

No tocante aos clientes que pagam mensalmente, 43% são classificados como *churn*. Enquanto que os clientes que pagam antecipado por um período de dois anos e os clientes que pagam antecipado por um período de um anos são classificados, respectivamente, em 3% e 11% como *churn*. Isso é visto na Figura 10.

Figura 10 – Tipo de contrato vs *Churn*

Diante destes dados, podemos identificar que há uma possível relação entre o cliente que paga mensalmente e o *churn*. Por isso, estes dados serão usados na análise preditiva.

Com relação ao tempo como cliente, notou-se que a base consiste em uma amostra de 32% de clientes com mais de 4 anos, 38% de clientes entre um ano e quatro anos e 31% de clientes com menos de um ano, conforme Tabela 9.

Tempo como cliente	N° Clientes	% N° de Clientes
01. Entre 1 e 6 meses	1.481	21%
02. Entre 7 e 12 meses	705	10%
03. Entre 1 e 2 anos	1.024	15%
04. Entre 3 e 4 anos	1.594	23%
05. Acima de 4 anos	2.239	32%
Total	7.043	100%

Tabela 9 – Tempo de cliente.

Como pode ser visto na Figura 11, quanto mais tempo como cliente, menor a representatividade de *churn*.

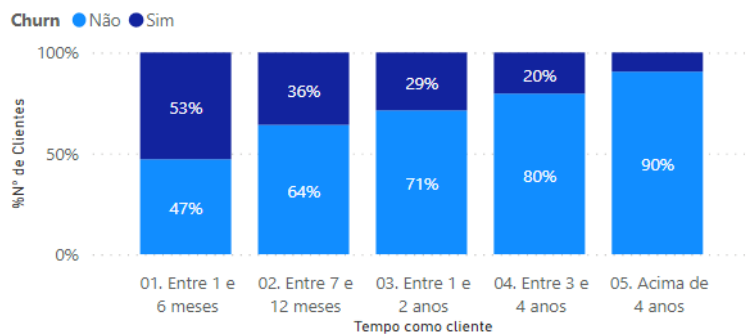


Figura 11 – Tempo de cliente vs *Churn*

Portanto, podemos concluir que há uma possível relação entre tempo como cliente e o *churn*. Por isso, estes dados serão usados na análise preditiva.

Com relação a forma de pagamento, conforme Tabela 10, identificou-se que a base consiste em uma amostra de 34% de clientes que pagam eletronicamente, 23% de clientes que pagam através de cheque, 22% de clientes que pagam por transferência bancária automática e 22% de clientes que pagam com o cartão de crédito em cobrança automática.

Forma de pagamento	N° Clientes	% N° de Clientes
Pagamento eletrônico	2.365	34%
Cheque	1.612	23%
Transferência bancária automática	1.544	22%
Cartão de crédito com cobrança automática	1.522	22%
Total	7.043	100%

Tabela 10 – Forma de pagamento.

Sobre os clientes que pagam eletronicamente, 45% são classificados como *churn*. Enquanto que os clientes com os demais formatos de pagamentos apresentam representatividades de 19% ou menos na classificação como *churn*. Isso é identificável na Figura 12.

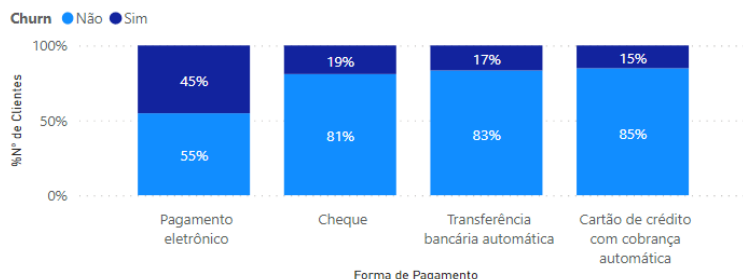


Figura 12 – Forma de pagamento vs *Churn*.

Portanto, podemos concluir que há uma possível relação entre o cliente que paga eletronicamente e o *churn*. Por isso, estes dados serão usados na análise preditiva.

3.2.1.4 Dados de satisfação

Os dados analisados são relacionados ao nota atribuída a satisfação por parte do cliente ao serviço e/ou produto usado. As notas concedidas são de 1 até 5, sendo que 1 é muito insatisfeito e 5 é muito satisfeito.

Conforme Tabela 11, com relação ao índice de satisfação, observou-se que a base consiste em uma amostra de 13% de clientes que classificam a satisfação em nota 1, 7% de clientes que classificam a satisfação em nota 2, 38% de clientes que classificam a satisfação em nota 3, 25% de clientes que classificam a satisfação em nota 4 e 22% de clientes que classificam a satisfação em nota 5.

Índice de satisfação	Nº Clientes	% Nº de Clientes
1	922	13%
2	518	7%
3	2.665	38%
4	1.789	25%
5	1.149	16%
Total	7.043	100%

Tabela 11 – Índice de satisfação.

De acordo com a Figura 13, conseguimos notar que os clientes com notas 1 ou 2 são em sua totalidade classificados como *churn*. Enquanto que os clientes com notas 4 ou 5 são em sua totalidade classificação como não *churn*.

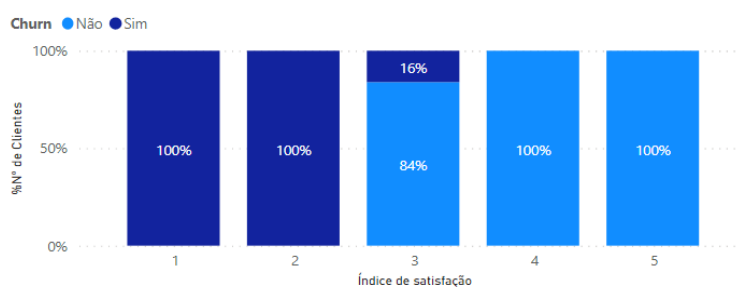


Figura 13 – Índice de satisfação vs *Churn*

Portanto, podemos concluir que há uma possível relação entre o cliente que classifica a satisfação com notas iguais ou abaixo de 2 e o *churn*. Por isso, estes dados serão usados na análise preditiva.

3.2.1.5 Dados selecionados

Conforme explicado na seção 3.1 e no início da seção 3.2, foram selecionados para a análise preditiva as seguintes categorias e dados:

- Dados de perfil: matrimônio.

- Dados relacionados a outros serviços contratados: *streaming*.
- Dados relacionados ao pagamento: tipo de contrato, tempo de contrato e modelo de pagamento.
- Dados de satisfação: índice de satisfação do cliente.

Esses foram os dados que apresentaram maior relação com o *churn* na análise descritiva.

3.2.2 Análise preditiva

Para iniciar a execução da análise preditiva, refez-se a limpeza da base de dados através da função *drop* do Python. O intuito é usar os dados definidos nos resultados obtidos na análise descritiva. Além disso, transformou-se os dados do tipo objetos em categóricos através da função *astype* para em seguida transformá-los em valores numéricos com a função *cat.codes*. Esta etapa é o pré-processamento dos dados e precede a manipulação dos dados para a predição.

Para executar o processo da predição, seguiu-se as seguintes sub-etapas de análise:

1. Divisão da base de dados entre base de treinamento e base de teste.
 - 1.1 Se necessário, balanceamento das bases de treinamento e teste.
2. Avaliação de acuracidade entre as técnicas de Regressão Logística e Árvore de Decisão.
3. Aplicação da técnica melhor avaliada.
4. Coleta e apresentação dos resultados.

Portanto, para criar as bases de dados para treinamento e teste, iniciou-se com a geração de duas bases de dados diferentes. Uma delas, a qual é denominada de X, é a cópia idêntica da base original, enquanto que a outra, chamada de Y, é uma versão sem os dados de *churn*. Para isso fez-se a cópia da base origem apenas removendo os dados de *churn* com a função *drop*.

Em seguida, aplicou-se a função *train_test_split*, que tem por objetivo gerar as bases de treinamento do algoritmo e de teste para a geração dos resultados. Usou-se como fonte as bases X e Y e em todas as funções foram usadas parâmetros padrões.

Após a divisão, identificou-se que o conjunto de dados é desequilibrado, chegando a proporção é de 1 para 3 entre base de teste e treinamento. Trabalhar com os dados dessa maneira impacta o desempenho da investigação. Portanto, para solucionar esse problema,

fez-se a balanceamento das bases através da sobreamostragem da classe minoritária com o uso das funções *smote* e *fit_resample*.

A sobreamostragem consiste em criar novos dados a partir da amostra já existente. O intuito desse processo é aumentar o volume de dados. A desvantagem deste formato é que o aumento quantitativo não é consequência de aumento qualitativo ou da diversidade da amostra. Isso pode tendenciar o resultado, porém, apesar disso, optou-se por considerar que isso não é um impeditivo para seguir esse caminho da análise. Até porque, o objetivo principal do projeto permite que não seja necessário rigidez nos critérios e requisitos da análise de dados, uma vez que o foco do Capítulo 3 é exemplificar como a análise de dados pode ser usada para a tomada de decisão nos negócios.

Logo depois, analisou-se duas técnicas de aprendizado de máquina em relação a acuracidade. Para isso, avaliou-se a Regressão Logística e a Árvore de Decisão e usou as funções *Pipeline*, *LogisticRegression*, *DecisionTreeClassifier* e *StandardScaler* combinadas entre si. Dentre as duas técnicas, a Regressão Logística foi a melhor avaliada com 0,9733 de acuracidade, enquanto que a Árvore de Decisão obteve 0,9699. A acuracidade é a medida que avalia a precisão e exatidão dos dados.

Conforme previsto, em sequência, aplicou-se a predição para Regressão Logística através da função *predict*. Dessa forma, obteve-se que 27% dos clientes da empresa fictícia tem tendências ao *churn*.

Além disso, com o objetivo de validar a análise descritiva, aplicou-se a função *corr* na base de dados original do início da análise preditiva. Essa função analisa a correlação dos entre dados, no caso entre os demais com o *churn*.

Essa medida varia de -1 até 1, sendo que quanto mais próximo de -1 ou 1 o resultados for, maior é a correlação entre os dados. Os valores negativos, mostram que a correlação é inversa, enquanto que os valores positivos mostram que a correlação é direta.

Como resultado, identificou-se os dados com maior correlação com o *churn*, os quais foram, em ordem decrescente:

1. Índice de satisfação, com correlação de -0,754649 com o *churn*.
2. Tipo de contrato, com correlação de -0.396713 com o *churn*.
3. Tempo de contrato, com correlação de -0.352229 com o *churn*.

Portanto, com uma correlação negativa, quanto maior os valores de índice de satisfação, tipo de contrato e tempo de contrato, menores são as chances de *churn* do cliente.

3.3 Análise dos resultados finais

Na análise descritiva, dos onze dados inicialmente analisados, identificou-se que seis deles tinham uma relação saudável de correlação com o *churn*. De forma que, como esperado, os dados de índice de satisfação foram os que se destacaram ao longo da análise. Isso vai de encontro com o afirmado pela literatura ao longo do projeto e, que por consequência, era a expectativa do resultado.

Os dados referentes ao índice de satisfação se sobressaíram de forma brusca. Isso porque, como visto no gráfico da Figura 13, os clientes insatisfeitos são todos classificados como *churn*, enquanto que os satisfeitos não tem nenhuma classificação de *churn*. Dessa forma, o alto índice de correlação inversa apenas reafirmou a influência entre os dados.

Outra expectativa com relação aos resultados era que houvesse uma identificação de correlação significativa entre o *churn* e os dados de uso de serviços e/ou produtos por parte do cliente. Contudo, o processo já chama a atenção na fase da análise descritiva, quando dos três tipos de dados dessa categoria (uso de múltiplas linhas telefônicas, *internet* e/ou *streaming*), apenas um (*streaming*) demonstrou relação com *churn* e foi selecionado para a análise preditiva.

Além disso, dentre os principais índices de correlação, o dado de *streaming* não aparece, de modo a trazer outra perspectiva para os resultados. Nesse cenário, é importante salientar que isso demonstra a importância da análise de dados para a tomada de decisão. Isso porque, se a decisão fosse tomada antes da análise, provavelmente alguma medida corretiva com relação ao uso de serviços e/ou produtos seriam feita, porém os dados demonstram que esse não é o caminho mais inteligente e assertivo.

Por outro lado, constatou-se que os dados com maior volume de impacto na rotatividade dos clientes da empresa fictícia analisada são relacionados ao pagamento. Dado que, dentre todas as categorias avaliadas, a de pagamento é a que mais apresentou dados com impactos facilmente visíveis na classificação do *churn* durante a análise descritiva. Essa também foi a única categoria que selecionou todos os dados da análise descritiva para serem avaliados na análise preditiva. Por fim, essa categoria também emplacou dois dos seus três dados dentre aqueles com maior índice de correlação. Portanto, é nítido a importância e influência dos dados de pagamento no *churn*.

Essa informação é uma surpresa e esse comportamento pode ser relacionado tanto à tipos de *churn* voluntários quanto involuntários. Por isso, para tomar decisões mais assertivas, seria recomendado investigar quais dos dois tipos de *churn* são os que mais estão acontecendo. Isso porque, no caso de inadimplência, ou *churn* voluntário acidental, precisamos agir diferente de casos em que o pagamento são apenas características em comum de um *cluster* de *churn* que na verdade está relacionado ao *churn* voluntário deliberado. Consequentemente, é mais sensato nesse caminho de investigação do que continuar com a

decisão inicial sem fundamentação que focaria apenas nos fatos relacionados ao uso de serviços e/ou produtos por parte do cliente.

Dessa forma, conseguimos atingir o segundo objetivo específico da análise, que era identificar quais são os principais comportamentos dos clientes que estão relacionados ao *churn*. Nesse caso a resposta é: o índice de satisfação, o tipo de contrato e o tempo de contrato. Pois, esses são os dados com maior índice de correlação com o *churn*.

A fim de complementar as análises descritivas, apoiar a tomada de decisão, e responder ao primeiro objetivo específico da análise, a análise preditiva afirmou que 27% dos clientes apresentam tendência ao *churn*. Esse é um valor alto, visto que no Capítulo 2 afirmamos que a porcentagem do mercado gira em torno de 2,2%. Logo, isso demonstra a urgência dessa situação para a empresa.

Nessa situação os dados são fictícios, mas se fosse uma empresa real isso seria um quadro crítico. De modo a demonstrar mais uma vez a importância de se utilizar a análise de dados para a tomada de decisão. Uma vez que, conseguimos identificar a criticidade e urgência do caso. De maneira a exigir que as medidas corretivas aconteçam de forma ágil e assertiva, e para isso é necessário se apoiar na análise de dados.

Em conclusão, diante dos resultados obtidos, a principal sugestão de ação inicial de curto prazo, sem aprofundar em outras análises de dados, seria incentivar e promover a venda de planos de contrato mais longos ao invés de planos de contrato curtos. Dessa forma, a empresa poderia demonstrar seus benefícios ao longo do contrato investindo em um relacionamento de qualidade com o consumidor para que o *churn* não acontecesse. Além de ter tempo para investir em uma análise mais direcionada para tomar outras decisões.

4 CONCLUSÃO

Conforme afirmou-se ao longo de todo o projeto, a transformação digital já é uma realidade há alguns anos, e os seus impactos foram mais acelerados com a chegada da pandemia do COVID-19. Dessa forma, a situação deixou claro a importância de ser ágil e assertivo nas empresas na hora de tomar decisões. Tanto é que, 3,5% das empresas brasileiras não conseguiram se adaptar e precisaram encerrar as suas atividades, enquanto que os 31% que se adaptaram de forma rápida estão sobrevivendo ao cenário.

Contudo, para ser ágil e assertivo, é necessário que as decisões sejam tomadas com base em dados, principalmente na atual situação que é um contexto digital. Essa afirmativa já é de conhecimento da literatura, porém, o projeto conseguiu exemplificar os impactos dessa assertividade e agilidade com o desenvolvimento da análise de rotatividade de clientes apresentada no Capítulo 3.

De acordo com a discussão exaustiva da seção 3.3, a análise desenvolvida tinha expectativas diferentes dos resultados obtidos. De modo que, ficou nítido a influência da análise de dados para a tomada de decisão, o que era o objetivo principal do projeto.

As conclusões após a análise direcionaram as medidas corretivas e preventivas para que fossem mais assertivas, inteligentes e ágeis. Além disso, a análise trouxe a perspectiva de criticidade e gravidade do cenário da empresa, o que não tinha sido mapeado anteriormente. Com isso, ficou evidente que, a análise de dados é importante tanto para a tomada de decisão quanto para expor cenários não previstos.

Se a empresa analisada, a qual é fictícia, fosse real e não tivesse como princípios usar a análise de dados para a tomada de decisão, provavelmente as decisões tomadas seriam diferentes da necessidade principal da empresa. Assim como, a velocidade de resolução não seria na medida necessária, pois não se teria conhecimento da gravidade da situação e, portanto, da urgência de priorização. Portanto, em um cenário de transformação digital e pandemia, isso é um problema grave de sobrevivência de mercado.

Por fim, o projeto conseguiu abordar como a análise de dados contribui para tomada de decisão diante do processo de transformação digital e contexto da COVID-19. Porém, além disso, o projeto também trouxe uma visão ampla da transformação digital e da implementação da análise de dados ao demonstrar como esse não um processo apenas de análise. Conforme apresentado, esse cenário também consiste em uma adaptação cultural, de estratégia empresarial e de gestão de pessoas, os quais quando abordamos dados não são tão facilmente correlacionados, mas que aqui é demonstrado a importância desse relacionamento.

REFERÊNCIAS

A, K. O.; NESTOR, D. J. Practical implementation of machine learning and predictive analytics in cellular network transactions in real time. In: . [S.l.: s.n.], 2018. Citado 2 vezes nas páginas 20 e 28.

ALMEIDA, R. M. C. *Classificação de churn no seguro automóvel*. Dissertação (Dissertação de Mestrado) — Universidade Nova de Lisboa, 2010. Citado 5 vezes nas páginas 15, 16, 17, 18 e 19.

ARAUJO, I. C. et al. Indústria 4.0 e seus impactos para o mercado de trabalho. *Brazilian Journal of Development*, v. 6, n. 4, p. 22326–22342, abr 2020. Disponível em: <<https://www.brazilianjournals.com/index.php/BRJD/article/view/9370/7915>>. Citado 3 vezes nas páginas 10, 13 e 14.

CARVALHO, G. G. de; TIOSSO, F.; REIS, H. M. Indústria 4.0: adoção de tecnologias como fator de competitividade. *Revista Interface Tecnológica*, v. 17, n. 2, p. 256–268, dez 2020. Disponível em: <<https://revista.fatectq.edu.br/index.php/interfacetecnologica/article/view/980>>. Citado 2 vezes nas páginas 10 e 13.

DALVI, P. K. et al. Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. In: IEEE. *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*. [S.l.], 2016. p. 1–4. Citado 2 vezes nas páginas 19 e 20.

DÍAZ, A.; ROWSHANKISH, K.; SALEH, T. *Why data culture matters*. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/why-data-culture-matters>, 2018. Citado 2 vezes nas páginas 21 e 22.

DIRK, V. d. P.; LARIVIÈRE, B. Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, v. 157, n. 1, p. 196–217, 2004. ISSN 0377-2217. Smooth and Nonsmooth Optimization. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0377221703000699>>. Citado 2 vezes nas páginas 6 e 18.

EVANS, J. R.; LINDNER, C. H. *Business Analytics: The Next Frontier for Decision Sciences*. 2012. Disponível em: <http://faculty.cbpp.uaa.alaska.edu/afef/business_analytics.htm>. Acesso em: 14 mai. 2021. Citado na página 22.

FERREIRA, A. L. L. *Modelo de identificação de churn rotacional nas comunicações móveis*. Dissertação (Dissertação de Mestrado) — Universidade do Porto, 2015. Citado 2 vezes nas páginas 16 e 17.

GEVERT, V. G. Análise de crédito bancário com o uso de modelos de regressão logística, redes neurais e support vector machine. *Universidade Federal do Paraná*, 2009. Citado na página 27.

IBM. *Guide to IBM Cognos Analytics Sample Data*. 2017. Disponível em: <<https://community.ibm.com/community/user/businessanalytics/participate/blogs/blogviewer?BlogKey=FF811D76-ABE0-4DF2-BCEA-917176FD72E4>>. Acesso em: 04 abr. 2021. Citado na página 26.

- IBM. *Telco customer churn (11.1.3+)*. 2019. Disponível em: <<https://community.ibm.com/community/user/businessanalytics/blogs/steven-macko/2019/07/11/telco-customer-churn-1113>>. Acesso em: 04 abr. 2021. Citado na página 25.
- IBM. *IBM Data and AI Accelerators*. 2020. Disponível em: <<https://community.ibm.com/accelerators/?context=analytics&industry=Telecommunications>>. Acesso em: 04 abr. 2021. Citado na página 25.
- JÚNIOR, D. S. G. et al. Efeitos da pandemia do covid-19 na transformação digital de pequenos negócios. *Revista De Engenharia E Pesquisa Aplicada*, v. 5, n. 4, p. 1–10, 2020. Citado 2 vezes nas páginas 10 e 14.
- LABORATÓRIO DE APRENDIZADO DE MÁQUINA EM FINANÇAS E ORGANIZAÇÕES. *Os Três Tipos de Aprendizado de Máquina*. 2017. Disponível em: <<https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/>>. Acesso em: 27 jun. 2019. Citado 2 vezes nas páginas 23 e 27.
- LOPES, J. E. F. *Satisfação, lealdade e retenção: um pre-experimento aplicado à telefonia móvel*. Dissertação (Dissertação de Mestrado) — Universidade de Uberlândia, 2007. Citado na página 20.
- LU, N. et al. A customer churn prediction model in telecom industry using boosting. *IEEE Transactions on Industrial Informatics*, IEEE, v. 10, n. 2, p. 1659–1665, 2012. Citado 2 vezes nas páginas 16 e 19.
- MARTINS, H. et al. *Transformações digitais no Brasil: insights sobre o nível de maturidade digital das empresas no país*. [S.l.], 2019. Citado 3 vezes nas páginas 11, 14 e 21.
- MISHRA, A.; REDDY, U. S. A novel approach for churn prediction using deep learning. In: IEEE. *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*. [S.l.], 2017. p. 1–4. Citado 2 vezes nas páginas 19 e 20.
- NETO, M. V. d. G. O processo crisp-dm aplicado na construção de uma solução para análise de risco de crédito. *Universidade Federal de Pernambuco*, 2018. Citado na página 27.
- REINSEL, D.; RYDNING, J.; GANTZ, J. F. *Worldwide Global DataSphere Forecast, 2020–2024: The COVID-19 Data Bump and the Future of Data Growth*. [S.l.], 2020. Citado na página 14.
- SACOMANO, J. et al. *Indústria 4.0*. Blucher, 2018. ISBN 9788521213710. Disponível em: <<https://books.google.com.br/books?id=PNCuDwAAQBAJ>>. Citado 5 vezes nas páginas 10, 11, 13, 15 e 21.
- SCHNEIDER, P. H. *Análise preditiva de Churn com ênfase em técnicas de Machine Learning: uma revisão*. Tese (Doutorado), 2016. Citado 2 vezes nas páginas 23 e 27.
- SNEADER, K.; STERNFELS, B. *Da sobrevivência à prosperidade: reinventando o retorno pós-COVID-19*. <https://www.mckinsey.com/featured-insights/future-of-work/from-surviving-to-thriving-reimagining-the-post-covid-19-return/pt-br>, 2020. Citado na página 21.

- SOUZA, E. M. d. M. Desafios da indústria 4.0 no contexto brasileiro: Uma revisão da literatura. *Revista ÎANDÉ*, v. 5, n. 1, p. 44–57, mar 2021. Citado na página 15.
- VILAÇA, M. L. C.; ARAUJO, E. V. F. d. *Tecnologia, Sociedade e Educação na Era Digital*. UNIGRANRIO, 2016. ISBN 9788588943698. Disponível em: <http://www.pgcl.uenf.br/arquivos/tecnologia,sociedadeeducacaonaeradigital_011120181554.pdf>. Citado 2 vezes nas páginas 10 e 13.
- XAVIER, J. E. M. *Influência da Maturidade de Business Intelligence Analytics sobre Usos dos Sistemas de Medição de Desempenho: Estudos de Caso*. Tese (Doutorado), 2018. Citado na página 22.