



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO



Previsão e Interpretação de Churn: Integrando Análise Causal e Aprendizado de Máquina para Estratégias de Retenção Efetivas

Gustavo Henrique Farias Bezerra

Orientador: Profa. Dra. Mariana Rodrigues De Almeida

Proposta de Dissertação de Mestrado
apresentada ao Programa de Pós-Graduação
em Engenharia de Produção da UFRN (área
de concentração: Pesquisa Operacional)
como parte dos requisitos para obtenção do
título de Mestre em Engenharia de Produção.

Natal, RN, Maio de 2024

Previsão e Interpretação de Churn: Integrando Análise Causal e Aprendizado de Máquina para Estratégias de Retenção Efetivas

Gustavo Henrique Farias Bezerra

Dissertação de Mestrado submetida em 27 de maio de 2024 pela banca examinadora composta pelos seguintes membros:

Prof. Dra. Mariana Rodrigues De Almeida (orientadora) PEP/UFRN

Prof. Dr. José Alfredo Ferreira Costa PEP/UFRN

Prof. Dr. Marcus Vinícius Dantas de Assunção IFRN

Prof. Dra. Adrina Cavalcante Marques UFPE

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Bezerra, Gustavo Henrique Farias.

Previsão e interpretação de churn: integrando análise causal e aprendizado de máquina para estratégias de retenção efetivas / Gustavo Henrique Farias Bezerra. - 2025.

93f.: il.

Dissertação (Mestrado) - Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Produção, Natal, 2025.

Orientação: Profa. Dra. Mariana Rodrigues De Almeida.

1. Churn - Dissertação. 2. CRM - Dissertação. 3. Análise Preditiva - Dissertação. 4. Aprendizado de Máquina - Dissertação. 5. Inferência Causal - Dissertação. I. Almeida, Profa. Dra. Mariana Rodrigues De. II. Título.

RN/UF/BCZM

CDU 658.5

*"Tudo está tão confuso e nos ares da
terra eu já me afoguei. Flash
Gordon amigo vem me salvar vem.
vem, vem, vem, vem, vem, vem",
Ronnie Von.*

*"O sol de tarde é tão quente. O sol
de tarde é tão quente. Onde está o
meu chapéu? Aonde está o meu
chapéu? Borboletas, Pirulitos. ",
Jupiter Maça (Flávio Basso).*

Agradecimentos

A minha orientadora que sempre esteve presente em todos os momentos da minha vida acadêmica e me guiou por toda a Agradeço profundamente a minha orientadora, Mariana Rodrigues De Almeida, pelo seu compromisso, orientação e dedicação incansável ao longo deste percurso acadêmico. Seu conhecimento, insights e conselhos foram inestimáveis para o desenvolvimento deste trabalho.

A minha família, cujo amor, apoio e encorajamento foram a força motriz por trás de todas as minhas conquistas. Agradeço aos meus pais, Fernando e Célia, pelo seu apoio incondicional, paciência e sacrifícios que fizeram ao longo dos anos para me proporcionar a melhor educação possível. As minhas queridas irmãs, Fernanda e Danielly, pelo incentivo constante e por ser minha fonte de inspiração. As minhas Tias Marli e Marlene pelo conforto e carinho. As minhas pequenas sobrinhas Maria, Luíza e Ester. E a Franciete minha tia do coração que sempre esteve comigo por toda a minha vida.

Expresso minha gratidão aos meus amigos, pela camaradagem, apoio mútuo e momentos de descontração que compartilhamos ao longo desta jornada. Suas palavras de encorajamento e compreensão tornaram os desafios mais leves e as vitórias mais significativas. Onde cito aqui a Quirino, Ederson, Fabrizio, Gabriel, João, Kali, Bruno, Douglas, Braga, Arthur, Cortez, Nadja, Lucas, Amanda, Brenno e tantos outros que sempre estiveram comigo.

Por último, mas não menos importante, agradeço aos professores José Alfredo e Marcus Vinícius que contribuíram para este trabalho, por meio de suas sugestões, críticas construtivas e pelo apoio.

Resumo

A globalização e o amplo uso da internet transformaram a relação entre consumidores e empresas, estabelecendo uma interação direta e ativa entre eles. Neste cenário, compreender o ciclo de vida do cliente é vital para manter a estabilidade operacional e financeira das organizações, com um foco aguçado em fatores que promovem a satisfação e a fidelidade do cliente. Diante da questão da *churn* – que reflete a perda de clientes – várias indústrias enfrentam desafios que impactam diretamente sua lucratividade e sustentabilidade. Assim, esta pesquisa tem como objetivo desenvolver uma ferramenta que aprimore a modelagem preditiva de *churn*, enriquecendo-a com análise causal para não apenas prever com mais precisão, mas oferecer interpretações claras das razões para a perda de clientes. Utilizando o conjunto de dados IBM Telco Customer churn, versão 11.1.3, como suporte empírico, o estudo busca identificar variáveis influentes no *churn* e avaliar estratégias eficazes de retenção. A abordagem metodológica inclui o uso de técnicas de aprendizado de máquina, como LGBM, combinadas com métodos avançados de análise causal, como *Double Robust machine learning* e modelagem de Efeitos de Tratamento Médio Condicional, CATE. Desenvolvendo ferramenta que auxilie na identificação dos fatores retenção de clientes, desde aspectos demográficos a natureza dos serviços prestados, analisando variáveis como tipo de contrato, gênero, idade, entre outros. Espera-se que os resultados validem as teorias de Wu et al. (2021) sobre previsão de *churn* e revelem perfis de clientes com maior propensão ao abandono, como exposto pelos autores de Rudd et al. (2021), contribuindo significativamente para o gerenciamento do relacionamento com o cliente e oferecendo dados estratégicos para o desenvolvimento de táticas de retenção mais assertivas.

Palavras-chave: *Churn*, CRM, Análise Preditiva, Aprendizado de Máquina, Inferência Causal.

Abstract

Globalization and the widespread use of the internet have transformed the relationship between consumers and companies, establishing a direct and active interaction between them. In this scenario, deeply understanding the customer lifecycle is vital to maintaining the operational and financial stability of organizations, with a keen focus on factors that promote customer satisfaction and loyalty. Faced with the issue of *churn* – which reflects customer loss – various industries face challenges that directly impact their profitability and sustainability. Thus, this research aims to develop a tool that enhances the predictive modeling of *churn*, enriching it with causal analysis to not only predict more accurately but also offer clear interpretations of the reasons for customer loss. Using the IBM Telco Customer churn dataset, version 11.1.3, as empirical support, the study seeks to identify influential variables in *churn* and evaluate effective retention strategies. The methodological approach includes the use of machine learning techniques, such as LGBM and decision trees, combined with advanced methods of causal analysis, such as Double Robust machine learning and Conditional Average Treatment Effects (CATE) modeling. The objective is the development of tools that assist in identifying the determining factors for customer defection, encompassing demographic aspects to the nature of services provided, analyzing variables such as contract type, gender, age, among others. It is expected that the results validate the theories of Wu et al. (2021) on *churn* prediction and unveil customer profiles with a higher propensity for abandonment, significantly contributing to customer relationship management and offering strategic data for the development of more assertive retention tactics.

Keywords: Churn, CRM, Predictive Analysis, Machine Learning, Causal Inference.

Sumário

Sumário	i
Lista de Figuras	iii
Lista de Tabelas	vii
Lista de Símbolos e Abreviaturas	ix
1 Introdução	1
1.1 Contextualização	1
1.2 Objetivos	4
1.2.1 Objetivo Geral	4
1.2.2 Objetivos Específicos	4
1.3 Justificativa	5
1.4 Estrutura da Dissertação	6
2 Referencial Teórico	9
2.1 Ciclo de vida do cliente	9
2.1.1 CRM - Gerenciamento da relação com o cliente	9
2.1.2 Churn	12
2.2 Aprendizado de Máquina	14
2.2.1 Regressão Logística	16
2.2.2 Árvores de Decisão	17
2.2.3 Modelos de <i>Gradiente Boosting</i>	20

2.2.4	Modelos de Redes Neurais Artificiais de Aprendizado Profundo - <i>Deep Learning</i>	21
2.3	Análise Causal	25
2.3.1	Efeitos de tratamento	25
2.3.2	Meta-aprendizado	28
2.4	Métricas Para Classificação	34
3	Trabalhos Relacionados	43
3.1	Literatura do Ciclo de Vida do Cliente	43
3.1.1	Previsão de Churn	43
3.1.2	Modelos de elevação - Análise Causal	47
4	Modelagem	49
4.1	Conhecimento de Dados, Negócio e Manipulação de dados	50
4.1.1	Dados e Negócio	50
4.1.2	Engenharia do modelo	58
4.1.3	Análise causal	61
4.2	Qualidade do modelo, Entrega e Manutenção	62
5	Resultados	65
5.1	Métricas de avaliação - Classificador	65
5.2	SHAP Values e Importância das Variáveis	71
5.3	Análise Causal	78
5.4	Análise de Resultados	82
6	Conclusão	85
	Referências bibliográficas	87

Lista de Figuras

1.1	Quantidade de trabalhos por ano. Fonte: Scopus.	6
1.2	Estrutura da dissertação.	8
2.1	Árvore de decisão de múltiplas variáveis. Fonte:	18
2.2	Arquitetura de Perceptron. Fonte:	21
2.3	Arquitetura de Perceptron. Fonte:	22
2.4	Arquitetura de Perceptron. Fonte:	24
2.5	Arquitetura do modelo de X-Learn. Fonte:	31
2.6	Arquitetura de modelo de DML. Fonte:	33
2.7	Gráfico de <i>lift</i> de um modelo de aprendizado de máquina. Fonte:	38
2.8	Gráfico de ganho cumulativo para um modelo de floresta aleatória. Fonte:	39
2.9	Gráfico de ganho cumulativo para um modelo de floresta aleatória. Fonte:	41
4.1	Distribuição de variáveis categóricas do perfil do cliente.	52
4.2	Distribuição de variáveis categóricas da utilização de serviços.	52
4.3	Distribuição de variáveis categóricas referentes a especificações contratuais.	53
4.4	Distribuição da quantidade de meses dos clientes na plataforma e sua propensão a saída.	54
4.5	Distribuição de cobranças totais e por mês realizadas.	55
4.6	Distribuição de cobranças totais e por mês realizadas.	56
4.7	Correlação de Pearson entre as variáveis.	56
4.8	Distribuição da variável ChurnLabel e Distribuição estratificada por tipo de contrato	57

4.9	Fluxograma do modelo de aprendizado de máquina.	59
4.10	DAG do processo de análise causal.	62
4.11	Exemplo do uso de SHAP em modelos de aprendizado de máquina.	64
5.1	Gráfico de relação entre precisão e sensibilidade para o modelo sem balanceamento.	66
5.2	Gráfico de relação entre precisão e sensibilidade para o modelo com balanceamento.	67
5.3	Gráfico da curva AUC-ROC para o modelo sem balanceamento.	67
5.4	Gráfico da curva AUC-ROC para o modelo com balanceamento.	67
5.5	Gráfico de "uplift" para o modelo sem balanceamento.	68
5.6	Gráfico de "uplift" para o modelo com balanceamento.	68
5.7	Gráfico de "wise uplift" para o modelo sem balanceamento.	69
5.8	Gráfico de "wise uplift" para o modelo com balanceamento.	69
5.9	Gráfico de ganhos acumulados para o modelo sem balanceamento.	69
5.10	Gráfico de ganhos acumulados para o modelo com balanceamento.	70
5.11	Estatística KS para o modelo sem balanceamento.	70
5.12	Estatística KS para o modelo com balanceamento.	71
5.13	Gráfico de dispersão de importância de variáveis do SHAP para o modelo sem balanceamento.	72
5.14	Gráfico de dispersão de importância de variáveis do SHAP para o modelo com balanceamento.	73
5.15	Distribuição <i>churn</i> por contrato.	74
5.16	Distribuição <i>churn</i> por idade do cliente.	74
5.17	Distribuição Churn por uso de streaming de filmes.	74
5.18	Distribuição <i>Churn</i> por contrato com base balanceada.	75
5.19	Distribuição <i>Churn</i> por idade do cliente com base balanceada.	75

5.20	Distribuição <i>Churn</i> por uso de " <i>streaming</i> de filmes com base balanceada.	76
5.21	Distribuição <i>Churn</i> por tempo na plataforma.	76
5.22	Análise de decisão do cliente por gráfico de cachoeira para modelo não balanceado.	77
5.23	Análise de decisão do cliente por gráfico de cachoeira para modelo balanceado.	78
5.24	Comparação de resultados para base balanceada.	82
5.25	Comparação de resultados para base balanceada.	82
5.26	Relação entre os resíduos e seus valores ajustados.	84

Lista de Tabelas

4.1	Variáveis categóricas da base de dados	50
4.2	Variáveis numéricas da base de dados	51
5.1	Métricas classificador LGBM - Modelo não balanceado.	66
5.2	Métricas classificador LGBM - Modelo balanceado.	66
5.3	Regressões por OLS ChurnLabel e Contrato.	80
5.4	Regressões por Random Forest e DML para churnResidual e ContratoResidual.	81

Lista de Símbolos e Abreviaturas

ATE Efeito de Tratamento Médio

CART "*Classification and Regression Tree*"

CATE Efeito de Tratamento Condicional Local

CRISP – ML(Q) "*Cross-Industry Standard Process for the development of machine learning applications with Quality assurance methodology*"

CRM "*Customer Relationship Management*"

DML "*Double Debiased machine learning*"

HTE Efeito de Tratamento Heterogêneo

LATE Efeito de Tratamento Médio Local

LGBM "*Light Gradient Boosting Machine*"

ML "*machine learning*"

RF "*Random Forest*"

ROC – AUC "*Receiver Operating Characteristic Area Under the Curve*"

Capítulo 1

Introdução

1.1 Contextualização

O *churn*, no âmbito empresarial, refere-se à taxa de rotatividade de clientes, sendo um indicador crucial para avaliar a saúde e sustentabilidade de uma base de clientes. Especialmente para empresas nos setores financeiro e tecnológico, o *churn* pode representar um desafio significativo, impactando diretamente a receita e a lucratividade. Compreender as razões por trás do *churn* é crucial para implementar estratégias eficazes de retenção de clientes.

A Gestão de Relacionamento com o Cliente (CRM) emerge como uma ferramenta essencial no combate ao *churn*, especialmente no setor de serviços. Os sistemas de CRM permitem que as empresas colem, armazenem e analisem dados detalhados sobre as interações dos clientes ao longo do tempo. Essa compreensão aprofundada do histórico de relacionamento proporciona insights valiosos sobre as preferências, padrões de compra e comportamentos dos clientes, permitindo a identificação antecipada de sinais de *churn*.

As causas do *churn* podem variar, desde a insatisfação com produtos ou serviços até mudanças nas circunstâncias dos clientes. A perda de clientes pode resultar de experiências negativas, concorrência acirrada, falta de personalização ou falhas na entrega de serviços. Empresas frequentemente utilizam análises de dados avançadas para identificar padrões e fatores preditivos que indicam a probabilidade de *churn*.

No contexto financeiro, a confiança é um componente central na relação entre cliente e empresa. A perda de confiança pode resultar em *churn*, e as instituições financeiras frequentemente implementam estratégias para garantir a segurança dos dados, a transparência nas transações e a conformidade regulatória.

Já no setor tecnológico, a inovação constante é essencial para reter a atenção dos clientes. Empresas que não conseguem acompanhar as mudanças tecnológicas correm o risco de perder clientes para concorrentes mais ágeis e adaptáveis. A análise preditiva desempenha um papel fundamental, permitindo que as empresas identifiquem padrões de comportamento do cliente que indicam a probabilidade de *churn*.

No contexto do varejo, uma das principais razões para o *churn* é a falta de fidelidade do cliente, muitas vezes impulsionada pela natureza competitiva do mercado varejista. Os consumidores têm uma variedade de opções e podem facilmente mudar de marca ou loja em busca de preços mais baixos, experiências de compra melhores ou produtos mais inovadores.

Assim, para todas as indústrias citadas, a compreensão das razões por trás do *churn* é crucial para implementar estratégias eficazes de retenção de clientes. O uso de *machine learning* desempenha um papel significativo contra o *churn*, oferecendo uma abordagem proativa e preditiva para identificar potenciais casos de perda de clientes. Algoritmos de aprendizado de máquina podem analisar grandes conjuntos de dados, identificando padrões sutis de comportamento do cliente que podem indicar a probabilidade de *churn*. Essa análise preditiva permite que as empresas tomem medidas preventivas, antecipando-se aos desafios e implementando estratégias personalizadas de retenção.

A inferência causal é outra ferramenta poderosa que aprimora a compreensão das relações de causa e efeito por trás do *churn*. Ao utilizar métodos avançados de inferência causal, as empresas podem identificar os fatores específicos que contribuem para a decisão do cliente de abandonar um serviço. Isso vai além da simples correlação de dados, permitindo uma compreensão mais profunda das causas subjacentes. Por exemplo, a infe-

rência causal pode revelar se a insatisfação do cliente está relacionada a um determinado aspecto do serviço ou se é influenciada por fatores externos.

A combinação de *machine learning* e inferência causal proporciona uma abordagem abrangente para compreender e reduzir o *churn*. Ao integrar algoritmos preditivos com análises causais e modelos de aprendizado de máquina causal, as indústrias podem não apenas prever a probabilidade de *churn*, mas também compreender as razões por trás dessas previsões. Isso permite a implementação de estratégias mais direcionadas e eficazes, abordando os problemas fundamentais que levam à perda de clientes. Assim, combinando trabalhos como os realizados por Wu et al. (2021) e Huang et al. (2012), que buscam prever o *churn* utilizando conceitos de aprendizado de máquina, a trabalhos como Verhelst et al. (2020), Rudd et al. (2021) e Devriendt et al. (2021) para gerar melhores ideias de como combater a perda de clientes.

Dessa forma, a presente dissertação usará o conceito combinado de aprendizado de máquina e análise causal para a construção de visões de combate ao *churn*, auxiliando a tomada de decisão entre entes gerenciais. Não obstante, serão utilizados conceitos de classificação e regressão para a previsão de quais possíveis clientes podem abandonar o produto, e será utilizado um "*framework*" de causalidade para a denotação de como certas variáveis podem alterar ou não uma possível política de prevenção de perdas de clientes.

A organização do presente trabalho é dividida pelas seguintes seções: 1- Introdução e objetivos, onde é conceituado a perda de clientes e os objetivos gerais e específicos a serem atingidos pelo estudo, 2- Referencial Teórico, onde serão apresentados os conceitos utilizados para a previsão e causalidade, 3- Trabalhos Relacionados, o capítulo onde são apresentados os estudos já realizados na área de *churn* desde previsão até planejamento estratégico, 4- Modelagem, onde é apresentado o conceito de pesquisa e as hipóteses a serem validadas do estudo, 5- Resultados, que vem a complementar o capítulo anterior ao revelar os "*outputs*" do conceito da pesquisa, 6- Conclusão, onde todo o trabalho é revisitado e são feitas revisões do resultado e do caminho pelo qual a dissertação trilhou.

1.2 Objetivos

Esta seção define o começo da arquitetura de pesquisa. A decisão de quais serão os objetivos da dissertação, seja ele geral ou específico, nasce a partir do conhecimento da literatura já produzida, dos axiomas e lemas definidos pelos autores de estudos de sustentação e o conhecimento do objeto analisado. Dessa forma, é o objetivo geral do presente trabalho e suas designações específicas para a construção e avaliação da dissertação.

1.2.1 Objetivo Geral

O objetivo geral do estudo é desenvolver ferramentas para o auxílio de tomada de decisão para o gerente/liderança da área de marketing de uma companhia de forma a buscar melhores políticas de retenção de clientes e expor melhor todo o processo de um ciclo de vida do usuário.

As ferramentas a serem construídas serão baseadas no conceito de aprendizado de máquina e buscarão sempre seguir o conceito padrão de mercado, como o CRISP-ML(Q), aliado a conceitos de análise causal e econometria. Serão estudados modelos de regressão e classificação para a criação de preditor de *churn* e será feita a verificação das possíveis causas da saída do usuário.

1.2.2 Objetivos Específicos

Para o teste das hipóteses formuladas e o cumprimento do objetivo geral, objetivos menores deverão ser atingidos. Sendo estes:

- Definir qual método utilizar para previsão do *churn* e garantir qualidade dos modelos utilizados;
- Definir qual método utilizar para análise causal do *churn* e garantir transparência no processo de descobrimento causal;

- Verificar a importância de cada variável perante os modelos de previsão e causalidade.

1.3 Justificativa

A crescente complexidade do ambiente empresarial, aliada à rápida evolução tecnológica, destaca a necessidade urgente de estratégias inovadoras para a gestão de clientes e a manutenção de bases sólidas. Nesse contexto, a previsão e análise de *churn*, ou a taxa de rotatividade de clientes, emergem como áreas cruciais para garantir a sustentabilidade e a competitividade das organizações.

Dos artigos encontrados presentes no repositório Scopus verificou-se que não há artigos que relacionem diretamente a previsão por aprendizado de máquina e inferência causal. Não obstante, na base web of science são recuperados 6 artigos. Combinando três palavras chaves (*churn*, *machine learning* e *causal inference*) para os dois repositórios, obtém-se a resultante de 321 artigos publicados em periódicos. É notável a crescente de documentos produzidos desde 2008 e Índia, China e Estados Unidos posam com o maior número de pesquisas realizadas. a Figura 1.1 expõem a quantidade de trabalhos produzidos ao longo dos anos.

Dessa forma, a aplicação de *machine learning* oferece a capacidade de desenvolver modelos preditivos precisos para antecipar o *churn* com base em dados transacionais e comportamentais com exposto por Huang et al. (2012) e Wu et al. (2021). No entanto, a limitação desses modelos reside na falta de capacidade de fornecer insights causais robustos. A integração da inferência causal permite identificar não apenas correlações, mas relações de causa e efeito, oferecendo uma compreensão mais profunda dos fatores que impulsionam o *churn* como propostos pelos autores de Devriendt et al. (2021), Rudd et al. (2021) e Verhelst et al. (2021a). Nesse contexto, o projeto desta dissertação encontra justificativa a partir do *gap* na literatura da combinação das duas abordagens e possui duas

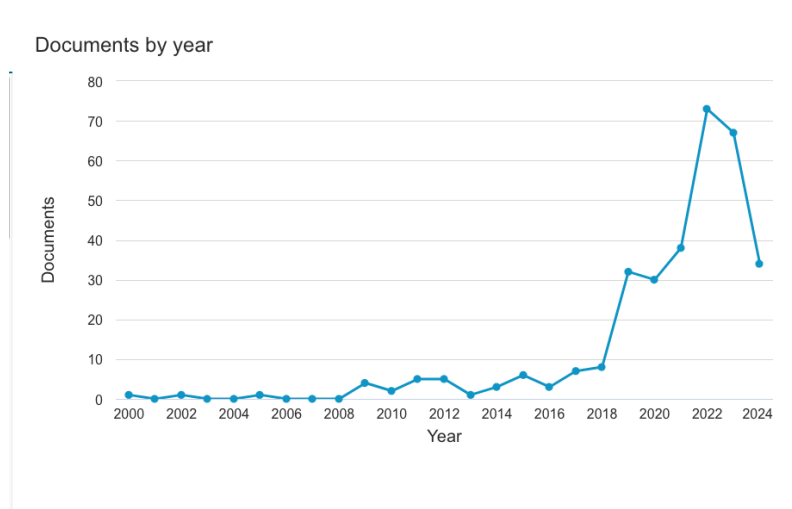


Figura 1.1: Quantidade de trabalhos por ano. Fonte: Scopus.

perspectivas, as quais são acadêmicas e socioeconômicas.

Do conceito acadêmico, a pesquisa propõe avançar o entendimento sobre a identificação da perda de clientes utilizando métodos de fronteira em aprendizado de máquina e análise causal expandindo e combinando os conceitos sedimentados por Rudd et al. (2021) e Wu et al. (2021).

Do conceito socioeconômico, o presente trabalho desenvolve soluções de fácil aplicação para problemas de decisão relacionados a retenção de clientes. Utilizando conceitos presentes em Lundberg e Lee (2017) aliado aos resultados encontrados nos trabalhos de Verhelst, como Verhelst et al. (2021a), tomadores de decisão e executivos podem obter melhores *insights* e definir melhores caminhos para soluções estratégicas no combate ao *churn*.

1.4 Estrutura da Dissertação

O capítulo 1 conceitua todo o estudo presente na pesquisa, divulga os objetivos pretendidos (geral específicos), justificativa e a estrutura do trabalho.

O capítulo 2 apresenta o referencial teórico para a pesquisa. A apresentação do con-

ceito de ciclo de vida do cliente, do gerenciamento da base de clientes e do *churn* são apresentados nesta seção. Adicionalmente, a apresentação geral de métodos de aprendizado de máquina, análise causal e metrificações de qualidade de modelos são definidos.

Da configuração do capítulo 3, são apresentados os trabalhos relacionados ao desenvolvimento da dissertação. Pesquisas em língua inglesa e portuguesa em relação a fronteira do conhecimento em previsão de perda de clientes e técnicas de aprendizado de máquina são expostas. De forma continuada, conceitos de análise causal aliados a conceituação de machine learning são apresentadas.

Do capítulo 4, A modelagem utilizada é apresentada. O conceito de do *framework* CRISP-ML(Q), a análise exploratória, a seleção de variáveis e sua engenharia, a escolha dos melhores modelos para a avaliação das hipóteses, a descrição do modelo causal e a avaliação por meio de SHAP são demonstrados.

Os resultados são apresentados a partir do capítulo 5. As métricas do previsor de *churn* bem como a apresentação da importância de cada variável para modelo são definidas nessa seção. Além disso, os resultados da modelagem causal também estrelam o capítulo.

Por fim, o capítulo 6 apresentam as considerações finais e conclusões da dissertação em relação a resolução das hipóteses propostas e de estudos futuros. A organização do presente trabalho é dividida conforme a Figura 1.2.

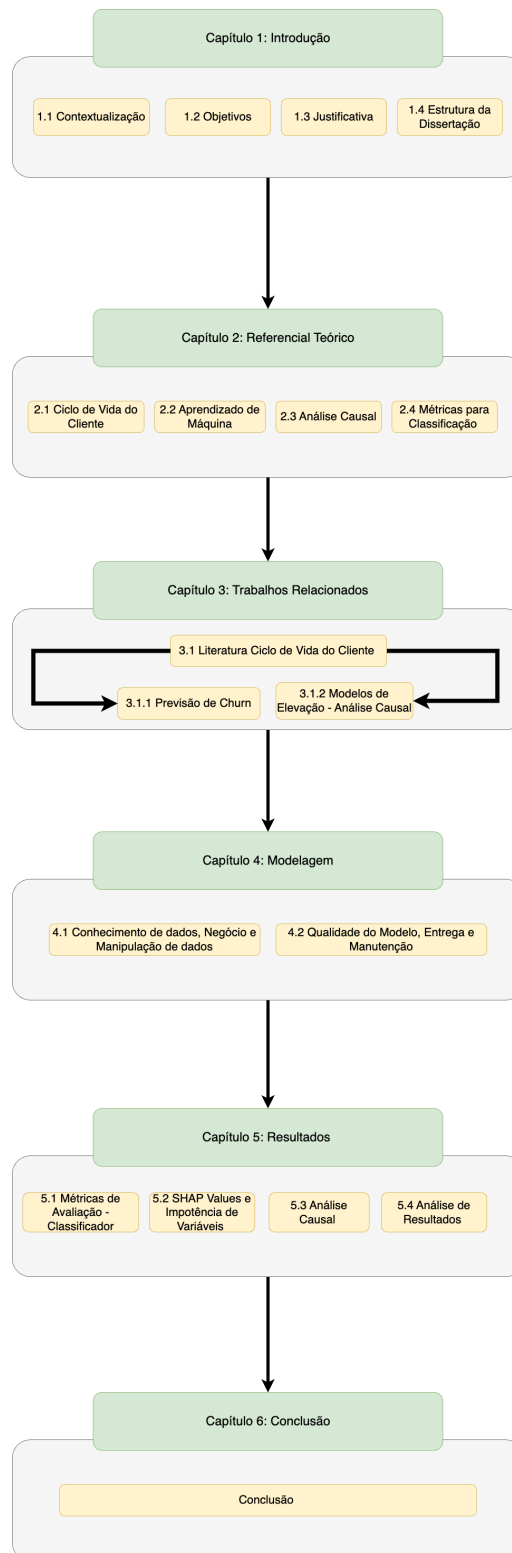


Figura 1.2: Estrutura da dissertação.

Capítulo 2

Referencial Teórico

2.1 Ciclo de vida do cliente

Neste capítulo, é apresentado o conceito de retenção de clientes na indústria, juntamente com as bases teóricas para sua detecção, análise, previsão e suas causas. Para tanto, exemplifica-se o conceito de ciclo de vida do cliente e aprofunda-se no conceito de perda desse consumidor, compreendendo o que é *churn* e seus fundamentos teóricos a partir do estudo realizado em Kurtz e Clow (1998). Dessa forma, é avaliada a taxa de possíveis usuários propensos a sair da plataforma, fornecendo material para a tomada de decisão dos gerenciadores da plataforma. Além disso, são apresentados modelos de aprendizado de máquina que serão empregados na previsão de uma possível perda de usuários e no efeito causal que políticas de retenção de clientes podem ter. Nesse contexto, faz-se a exposição de modelos clássicos de *machine learning*, como árvores de decisão, florestas aleatórias e modelos voltados ao gradiente descendente como exemplificados em Schneider (2016), bem como *frameworks* causais para a valoração de efeitos causais heterogêneos presentes em Facure (2020).

2.1.1 CRM - Gerenciamento da relação com o cliente

O ciclo de vida do usuário da plataforma é objeto de estudo em áreas correlatas de inteligência de mercado, marketing e comercial. Dessa forma, define-se o que é entendido

como **CRM**, do inglês "*Customer Relationship Management*", que lida com todo o ciclo de vida do cliente. Utilizando as fundamentações teóricas de Swift (2001) e Kurtz e Clow (1998), são definidas cinco grandes fases na vida útil do usuário em relação ao produto: aquisição, cadastro, consumo, retenção e recuperação.

Aquisição

Nesta etapa, a plataforma busca identificar e atrair clientes em potencial com base no plano de uso do produto. Isso pode ser realizado por meio de estratégias de marketing, sejam elas clássicas ou digitais. Anderson (2008) disserta que a utilização do conceito de SEO, do inglês "*Search Engine Optimization*", impulsiona a obtenção de novos clientes de maneiras focadas no cliente com maior valor inerente à vida útil do produto. Por conseguinte, gerando maior prospecção de futuros clientes.

Cadastro

Como exposto por Kurtz e Clow, após a aquisição, o cliente precisa ser cadastrado, ou seja, integrado à plataforma ou ao software de CRM. Isso inclui a realização de procedimentos como a abertura de conta, a coleta de informações, configuração de perfil e definições legais. Neste ponto, o usuário já é parte integrante da população a ser estudada e pode ser submetido a testes/experimentos de novas políticas de marketing. Os autores de Kurtz e Clow (1998) argumentam que o uso de boas práticas e *softwares* ligados a gerências de clientes mantém a estrutura do ciclo de vida.

Consumo

Nesta etapa, o cliente começa a utilizar os serviços da plataforma como um todo. Nesta etapa, o cliente começa a utilizar os serviços da plataforma como um todo, como exposto em Swift (2001). Neste momento, o cliente passa a fornecer informações ao produto, desde conceitos de utilização para a melhoria de sua experiência até informações

de sinais de preço e demanda favorecendo toda a esteira de produção. As informações fornecidas pelos clientes expõem o panorama do produto e são utilizadas para tomada de decisão. Neste ponto, pode-se dividir o futuro do ciclo de vida em dois: Retenção, lealdade do usuário perante o produto, e recuperação, tentativa de conseguir reter clientes próximos ao abandono do produto ou que já deixaram de utilizá-lo, como argumentado em Schneider (2016). O ponto de estudo deste trabalho refere-se a esses últimos pontos -Retenção e Recuperação. Não obstante, o conceito de previsão e causa de *churn* são pontos válidos para o *decision maker* construir novas abordagens à companhia e ao produto.

Retenção

O objetivo desta etapa é manter o cliente satisfeito e fidelizado, evitando possíveis desligamentos. Isso pode ser alcançado por meio de ofertas personalizadas, atendimento de qualidade e outros programas de relacionamento compreendidos pela equipe de relacionamento com o cliente. Kurtz e Clow (1998) alertam que o custo para a aquisição de novos clientes pode chegar a ser 5 vezes maior do que o de retenção de clientes ativos do produto. Portanto, o bom relacionamento e a previsão de possíveis contratemplos ajudam a saúde financeira e gerencial do produto.

Recuperação

A etapa é voltada para o retorno dos usuários já desligados da base do produto. O objetivo é identificar as causas da desistência e buscar maneiras de recuperar o cliente, sendo as utilizações de modelagens causais importantes nesse ponto. Rudd et al. (2021) apresentam um estudo referente à retenção de clientes a partir do entendimento das causas da saída dos clientes. Utilizando redes neurais artificiais profundas, os autores definem que variáveis como aumento do preço e idade são fatores determinantes para isso.

2.1.2 Churn

Os conceitos de *churn* e CRM (Customer Relationship Management) estão intrinsecamente relacionados no contexto do ciclo de vida do cliente em uma plataforma. O *churn*, como axioma exposto por Kurtz e Clow (1998), refere-se à taxa de perda de clientes, ou seja, aqueles que deixam de utilizar os serviços ou produtos oferecidos pela plataforma. Esse fenômeno pode ocorrer em várias fases do ciclo de vida do cliente, desde a aquisição até o uso contínuo.

Por sua vez, o CRM lida com a gestão do relacionamento com o cliente ao longo de todo o ciclo de vida, como explanado pelo autor de Schneider (2016). No contexto apresentado, o CRM é fundamental para as etapas de aquisição, cadastro, consumo, retenção e recuperação. Durante a aquisição, estratégias de marketing e técnicas como SEO são aplicadas para atrair novos clientes, sendo o CRM responsável pela integração desses clientes à plataforma.

Na fase de consumo, o cliente fornece informações valiosas que alimentam o CRM, contribuindo para a análise do comportamento do usuário e para a tomada de decisões. Quando o *churn* é abordado, o CRM desempenha um papel crucial na retenção de clientes ativos, proporcionando um bom relacionamento, ofertas personalizadas e antecipando possíveis desafios que podem levar à desistência. Contudo, é válido usar os conceitos propostos por em Schneider (2016) e Rudd et al. (2021) que dialogam sobre o correto uso de estratégias de relacionamento com o cliente. De forma continuada, os autores desses estudos apresentam que políticas devem ser testadas e seus resultados avaliados para o mantimento correto do uso de ferramentas de CRM.

Além disso, na fase de recuperação, o CRM pode ser empregado para identificar as causas do *churn* e desenvolver estratégias para reconquistar clientes desligados, como bem expostas na revisão apresentada em Schneider (2016). Modelagens causais podem ser utilizadas para entender as variáveis determinantes do *churn*, como aumento de preço e idade, permitindo a criação de abordagens mais eficazes para evitar a perda de clientes

como utilizadas em Rudd et al. (2021) e Verhelst et al. (2020). Portanto, o CRM atua como uma ferramenta estratégica em todas as fases do ciclo de vida do cliente, desempenhando um papel crucial na gestão do *churn* e na manutenção de relacionamentos duradouros com os usuários da plataforma.

Utilizando o conhecimento do ciclo de vida mencionado nos parágrafos anteriores, pode-se definir o *churn* como a taxa de usuários que deixam de utilizar o produto ao qual estão cadastrados. Dessa forma, compreende-se que a taxa de saída de clientes, *churn rate*, é descrita como a relação percentual entre o número de clientes ativos e o número total de clientes que saíram da plataforma. A Equação 2.1 apresenta de forma sucinta essa relação, onde C representa os clientes e é obtida por meio do exposto pelos autores de Kurtz e Clow (1998).

$$ChurnRate = \frac{(C_{ativos} + C_{novos} - C_{desativados})}{C_{ativos}} \quad (2.1)$$

Nesse contexto, a utilização de ferramentas que possam auxiliar o entendimento e a diminuição dessa taxa são entes necessários a equipe que lidará com o ciclo de vida do usuário. Estudos como os apresentados pelos autores de Ghorbani e Taghiyareh (2009) dialogam e demonstram que a perda de clientes é entendida por dois fatores. Primeiro, a migração do usuário para produto de empresa concorrente devido a insatisfação diminuindo, assim, a base de clientes. Segundo, a diminuição do resultado financeiro dado a saída de clientes pagantes da plataforma.

Eva Ascarza argumenta que a retenção de clientes de menor risco tende a ser melhor financeiramente para o produto do que focar em consumidores de alto valor, mas voláteis. Em Ascarza (2018), ela evidencia de forma analítica este argumento e demonstra que políticas de retenção tendem a apresentar resultados heterogêneos. No estudo Veloso (2012), o autor apresenta formas de diagnosticar as causas de perda de clientes e prevê-las em in-

dústrias têxteis. Ele utiliza modelos de aprendizado de máquina, como florestas aleatórias e "Support Vector Machines" para definir um melhor preditor de *churn* e compará-lo a modelos clássicos.

De forma contínua, a expansão de metodologias para a prevenção do *churn* encontra na inteligência artificial, essencialmente no aprendizado de máquina, o motor que entrega melhores previsões e entendimentos mais aprofundados. Em Schneider (2016), o autor apresenta uma revisão sistemática de trabalhos que relacionam o conceito de previsão de *churn* ao *machine learning*, encontrando cerca de 80 artigos. Assim, o autor expõe o crescente interesse nas áreas e a crescente usabilidade de metodologias de analítica avançada.

O trabalho Rudd et al. (2021) apresenta uma comparação de métodos de aprendizado de máquina, incluindo um modelo de redes neurais artificiais profundas, para demonstrar os melhores resultados ao utilizar essas técnicas, em comparação com as análises comumente realizadas pelas áreas de relacionamento com o consumidor.

Ampliando os conceitos de previsão de *churn*, os autores de Pinheiro e Cavique (2022) apresentam, conceptualmente e por meio de um estudo de caso, as razões pelas quais os clientes de provedores de serviços de telecomunicações tendem a deixar a companhia. Os autores demonstram que consumidores com pagamentos anuais têm uma taxa de desistência menor do que outras formas de contrato. Combinando o conceito de *machine learning* e análise causal, os trabalhos Rudd et al. (2021) e Verhelst et al. (2021b) diagnosticam o conceito de CATE, Efeito de Tratamento Médio Condicional, e a heterogeneidade do resultado do tratamento nos clientes que receberam a nova política.

2.2 Aprendizado de Máquina

No contexto do *churn*, o termo *machine learning* refere-se à aplicação de algoritmos e técnicas computacionais de otimização e inferência estatística para prever ou identificar

a probabilidade de um cliente ou usuário interromper o uso de um produto ou serviço. *Churn*, ou "perda de clientes", é uma métrica importante para empresas, pois representa a taxa de perda de clientes ao longo do tempo. Dessa forma, a utilização do conceito de *churn rate* para a métrica dessa perda é valorada, como bem argumentado por Schneider (2016).

Através da análise de dados históricos de clientes, seus padrões de comportamento, suas transações, suas interações e outros dados relevantes, o previsor pode obter melhores resultados e expandir a análise posterior. Os algoritmos de *machine learning* podem ser treinados usando esses dados para identificar padrões e construir modelos preditivos que ajudam a prever quais clientes têm maior probabilidade de deixar de usar um produto ou serviço. Existem vários tipos de modelos de abordagem no aprendizado de máquina que podem ser usados para prever o *churn*; assim, utilizando o estudo de Schneider (2016) como base, tais padrões são definidos como:

- **Modelos de classificação:** Esses modelos são utilizados para prever se um cliente irá ou não abandonar uma empresa, por meio da análise de padrões de clientes que já saíram da base de ativos.
- **Modelos de regressão:** Esses modelos são empregados para prever se um cliente irá ou não abandonar uma empresa. Utilizando conceitos de regressão, os modelos tendem a encontrar a probabilidade de saída a partir de variáveis que descrevam o *churn*.
- **Modelos de agrupamento:** Esses modelos são utilizados para identificar grupos de clientes que têm probabilidade semelhante de abandonar uma empresa. Agrupando clientes com características semelhantes, a tomada de decisão se beneficia de padrões comuns a grandes grupos.

Para o presente trabalho, foi decidido utilizar o conceito de modelagem a partir de classificação, seguindo o conceito mais utilizado na literatura e que melhor prevê e agrega

as possíveis saídas, como exemplificado pelos autores de Wu et al. (2021), Rudd et al. (2021) e Schneider (2016). Dessa forma, conceitos de aprendizado de máquina comumente aplicados ao *churn* incluem: Regressão Logística, Árvores de Decisão, Florestas Aleatórias, *Gradient Boosting* e Redes Neurais de Aprendizado Profundo (*Deep Learning*). Os conceitos de explicabilidade, acurácia e precisão devem guiar a produção da análise.

Ao empregar essas técnicas, as empresas podem reconhecer clientes em risco de *churn* e implementar estratégias específicas para reter esses clientes. Isso pode abranger ofertas personalizadas, programas de fidelidade, melhorias nos produtos ou serviços, entre outras ações voltadas a manter os clientes satisfeitos e engajados. Como mencionado anteriormente, o aprendizado de máquina desempenha um papel fundamental na retenção de clientes e nas possíveis recuperações.

2.2.1 Regressão Logística

A regressão logística, a partir do que é citado por Wooldridge (2016), modela a relação entre uma variável dependente binária (Y) e uma ou mais variáveis independentes ($X_1 + \dots + X_n$). No contexto da previsão, a variável dependente geralmente é binária, assumindo valores 0 ou 1, representando se o evento aconteceu ou não. A Equação 2.2 expressa essa relação.

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (2.2)$$

A interpretação dos coeficientes na regressão logística é fundamental para entender como cada variável independente contribui para a probabilidade de ocorrência do evento de interesse, que, no caso da previsão de *churn*, é a probabilidade de um cliente deixar de utilizar um produto ou serviço. O conceito de corte para tais probabilidades é fun-

damentado pela formula de averiguação. Nesse contexto, Wooldridge (2016) apresenta conceitos de distribuições normais de probabilidade, probit, e distribuições logarítmicas, logit.

Assim, o conceito das variáveis da regressão pode ser entendido a partir do Intercepto (β_0) representa o log-odds de ocorrência do evento quando todas as variáveis independentes são zero, o lugar onde todas as variáveis de controle tem efeito nulo. Assim, é o resultado padrão de ocorrência do *churn* quando todas as variáveis explicativas têm um valor de zero.

As variáveis seguintes apresentam sua função de distribuição de probabilidade e seu efeito na variável alvo por meio dos parâmetros da regressão (β_1, \dots, β_n), representando a mudança na probabilidade de ocorrência do evento para uma unidade de mudança em x_n , mantendo todas as outras variáveis constantes. Em resumo, é o efeito da variável alvo. Dessa forma, a análise dos resultados indicará quais variáveis são mais proeminentes na desistência do cliente, como indicado na teoria exemplificada no capítulo 17 de Wooldridge (2016).

2.2.2 Árvores de Decisão

Árvores de decisão são uma técnica de modelagem preditiva frequentemente utilizada para a previsão de churn. Essa abordagem é particularmente útil para representar decisões sequenciais em um formato de árvore, onde cada nó interno representa uma escolha baseada em uma característica específica, e as folhas da árvore representam as previsões. Como exemplificado pelos autores de Izza et al. (2020), os modelos de árvore são definidos como grafos cujos nós, a interseção dos galhos, representam uma decisão baseada em uma variável x_n . Dessa forma, cada ramificação a partir de um nó interno representa uma escolha ou teste condicional, e cada folha representa o resultado ou a previsão. A Figura 2.1 exemplifica como as árvores de decisão funcionam.

As árvores de decisão visam dividir os dados de treinamento de uma maneira que

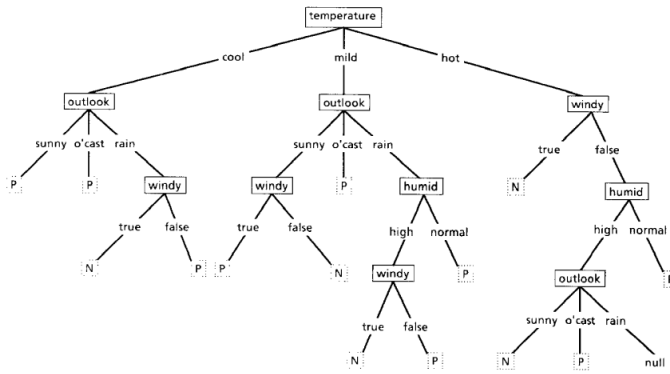


Figura 2.1: Árvore de decisão de múltiplas variáveis. Fonte: Quinlan (1986), p. 88.

reduza a incerteza sobre a variável de destino, reduzindo a entropia como Izza, Ignatiev e Silva demonstram em seu estudo já introduzido, o *churn* no caso deste estudo; entende-se entropia como medida de incerteza entre variáveis da base de dados. Por conseguinte, o ganho de informação é usado para determinar a melhor divisão em um determinado nó. Como evidenciado em Schneider (2016), a utilização de *decision trees* para modelos de classificação pode ser expressada em modelos CART, "*Classification and Regression Tree*". Os modelos CART fazem uso de modelagem de entropia por Gini, avaliando quantas vezes uma classe é escolhida aleatoriamente, retornando assim ramificações com menor impureza. A Equação 2.3, advinda do trabalho de Schneider (2016), demonstra o cálculo de entropia.

$$Entropia(D) = - \sum_{i=1}^c p_i \log_2(p_i) \quad (2.3)$$

As árvores de decisão oferecem uma abordagem intuitiva e eficaz para a previsão de churn, e suas decisões são facilmente compreendidas. No entanto, é importante considerar a tendência de sobreajuste, ou *overfit*, em árvores profundas, o que pode ser gerenciado ajustando a profundidade máxima da árvore ou utilizando técnicas como *pruning*.

Florestas Aleatórias

As florestas aleatórias, também conhecidas como "*Random Forests*", são uma extensão das árvores de decisão. O conceito definido por Cutler et al. (2012) para florestas aleatórias dialoga que enquanto as árvores de decisão individuais podem ser propensas ao *overfit*, ajuste excessivo, aos dados de treinamento, as florestas aleatórias buscam mitigar esse problema, ao mesmo tempo em que mantêm as vantagens da flexibilidade e interpretabilidade associadas às árvores de decisão, no ponto de expandir a previsão a partir do balanceamento por votação da previsão por diversas árvores em conjunto. As florestas são utilizadas como preditor de *churn* pelos autores de Rudd et al. (2021) e Wu et al. (2021) para o problema de previsão e inferência ao *churn*.

Por conseguinte, cada árvore é treinada em uma amostra aleatória do conjunto de dados, por meio de uma amostragem por *bootstrap*. A amostragem de *bootstrap* envolve a seleção aleatória de observações com substituição, onde observações podem aparecer várias vezes, enquanto outras podem não aparecer. Desta forma, há menor chance de correlação entre os nós e resultados das árvores. A previsão final é obtida agregando as previsões de cada árvore, onde em problemas de classificação, a maioria das votações é usada. Há a possibilidade de retorno da probabilidade de uma observação por meio dos resultados.

O ganho de informação em florestas aleatórias é calculado considerando a média ou a votação por maioria dos ganhos de informação de cada árvore. Presente na Equação 2.4, definida em Cutler et al. (2012), onde T indica o número de árvores na floresta e $\text{Ganho}_t(D, A)$ é o ganho de informação da árvore t .

$$\text{GanhoFloresta}(D, A) = \sum_{t=1}^T \text{Ganho}_t(D, A) \quad (2.4)$$

2.2.3 Modelos de *Gradiente Boosting*

Modelos de *Gradient Boosting* se destaca em previsões de conjuntos de dados complexos e grandes. Como definido pelos autores de Natekin e Knoll (2013), o modelo é construído sobre o princípio de treinar modelos fracos, geralmente árvores de decisão rasas, e combiná-los para formar um modelo mais robusto e preciso. Os modelos de função gradiente descendente funcionam de forma sequencial, treinando uma série de modelos fracos, chamados "estimadores fracos", e ajustando cada modelo para corrigir os erros dos modelos anteriores. A revisão apresentada por Schneider (2016) explora tal metodologia como a mais utilizada entre os trabalhos analisados e com resultados positivos.

Geralmente, os modelos fracos são árvores de decisão rasas ("*stumps*"), essas árvores têm profundidade limitada, o que as torna modelos simples e evita *overfit*. Logo, a cada iteração, o *Gradient Boosting* ajusta o modelo para os resíduos do estudo anterior de forma a tentar corrigir os erros das iterações anteriores. O processo segue atualizando seus *stumps* de forma oposta ao gradiente da função perda, "*loss function*", em relação as previsões realizadas. A fórmula geral para o ajuste do modelo no *Gradient Boosting* é obtida por meio da formulação de Natekin e Knoll (2013) e está presente na Equação 2.5.

$$F_t(x) = F_{t-1}(x) + \eta \cdot h_t(x) \quad (2.5)$$

onde:

- $F_t(x)$, representa o modelo atualizado na iteração t ;
- $F_{t-1}(x)$, modelo da iteração anterior;
- η , taxa de aprendizado;
- $h_t(x)$, estimador fraco na iteração t .

A previsão final é obtida somando as previsões de todos os modelos, exposta na Equação 2.6.

$$\text{Previsão Final}(x) = F_T(x) = \sum_{t=1}^T \eta \cdot h_t(x) \quad (2.6)$$

2.2.4 Modelos de Redes Neurais Artificiais de Aprendizado Profundo - *Deep Learning*

Deep Learning é a abordagem de aprendizado de máquina que envolve o treinamento de modelos com múltiplas camadas, conhecidos como redes neurais profundas, para aprender representações complexas e hierárquicas dos dados. Embora a aplicação de *Deep Learning* para previsão de *churn* possa ser poderosa, é importante considerar a complexidade do modelo em relação à quantidade e qualidade dos dados disponíveis. As redes neurais artificiais são a base do *Deep Learning*. Do conceituado em Belcak e Wattenhofer (2023), elas consistem em neurônios organizados em camadas, incluindo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída, composta por diversos neurônios em diversas camadas. Cada recurso dessas redes neurais necessita de uma função de ativação não linear para o início do processo de aprendizagem. Um exemplo de redes neurais artificiais é visto na Figura 2.2.

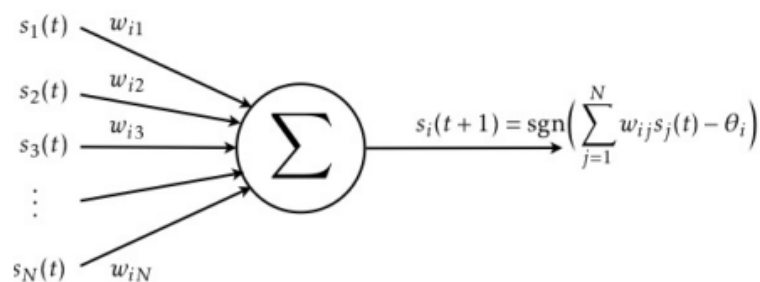


Figura 2.2: Arquitetura de Perceptron. Fonte: Mehlig (2019), p. 8.

Arquiteturas mais profundas, como Redes Neurais Profundas (DNNs), Redes Neurais Convolucionais (CNNs) e Redes Neurais Recorrentes (RNNs), têm a capacidade de

aprender padrões mais complexos e capturar dependências temporais. Em resumo, esse conceito pode ser simplificado em dois tipos de rede: "*Feedforward*" e "*Backpropagation*".

Propagação Direta - *Feedforward*

A propagação direta, também conhecida como *Feedforward*, é uma fase essencial em redes neurais artificiais (RNAs). A partir do que foi estudado pelos autores de Belcak e Wattenhofer (2023), nessa fase, as entradas são passadas pela rede, camada por camada, até chegar à camada de saída, sem que haja retroalimentação. A propagação direta é o processo de gerar previsões ou saídas com base nas entradas fornecidas à rede. As etapas iniciais envolvem a imputação de dados (X) e a associação de pesos (W) e vieses (b). Cada conexão entre neurônios na rede possui um peso associado (W) e um viés (b), sendo a soma ponderada das entradas associadas à rede neural. Essa soma é calculada pela Equação 2.7, onde Z representa a função de ativação do neurônio. O conceito de redes *Feedforward* é ilustrado na figura 2.3.

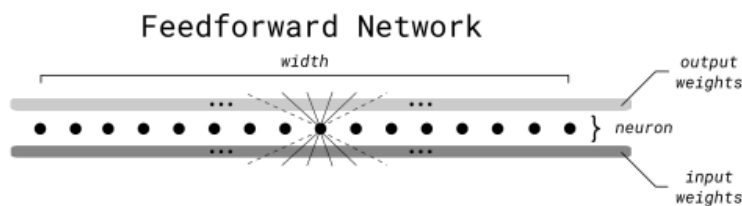


Figura 2.3: Arquitetura de Perceptron. Fonte: Belcak e Wattenhofer (2023), p.1.

A soma ponderada é submetida a uma função de ativação para introduzir não linearidades e permitir que a rede aprenda relações complexas. As saídas da camada atual tornam-se as entradas para a próxima camada, repetindo-se esse processo até chegar à camada de saída. O termo "aprendizado profundo" provém do fato de que as redes neurais podem possuir de 2 a n camadas ocultas.

$$Z = W \cdot X + B \quad (2.7)$$

Por fim, a função *Feedforward* de cada camada l pode ser descrita como em 2.8.

$$Y^l = a(W^l X^l + b^l) \quad (2.8)$$

Pelos quais:

- Y^l , saída da camada l ;
- a , função de ativação;
- W^l , matriz de pesos da camada l ;
- X^l , vetor de entradas da camada l ;
- b^l , vetor de viés da camada l .

Retropropagação - *Backpropagation*

A retropropagação, *Backpropagation*, é o algoritmo fundamental para treinar redes neurais, permitindo a atualização dos pesos da rede para minimizar a função de perda. Esse processo ocorre iterativamente durante o treinamento da rede e é conceituada por meio dos resultados de Rumelhart et al. (1986). Inicialmente, a rede realiza a propagação direta (*Feedforward*), gerando previsões com base nas entradas fornecidas. Em seguida, a função de perda avalia a discrepância entre as previsões da rede e os valores reais. O algoritmo de *Backpropagation* calcula os gradientes da função de perda em relação aos pesos da rede, propagando esses gradientes de volta (daí o nome "retropropagação") e atualizando os pesos para minimizar a discrepância.

Este processo de ajuste iterativo é fundamental para o aprendizado eficaz da rede

neural, pois permite que a rede se adapte aos padrões nos dados de treinamento. A figura 2.4 ilustra esse conceito, destacando a retropropagação como parte integrante do ciclo de treinamento de uma rede neural.

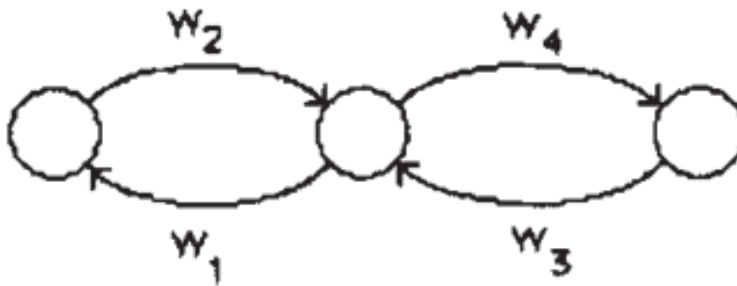


Figura 2.4: Arquitetura de Perceptron. Fonte: Rumelhart et al. (1986), p.535.

A fórmula geral para a atualização dos pesos durante a retropropagação é encontrada nos estudos dos autores de Rumelhart et al. (1986) e é expressa por meio do conceito de gradiente descendente, conceito de otimização usado para ajustar os pesos de uma rede neural de forma a encontrar os melhores máximos locais. Em resumo, os gradientes dos pesos são calculados em relação à função de perda. A regra da cadeia é utilizada para calcular esses gradientes, decompondo a derivada de uma função composta em derivadas de funções individuais. A atualização dos pesos é realizada de forma a minimizar a função objetivo, aproximando os resultados ao ótimo desejado. As equações 2.9 e 2.10 ilustram o processo de atualização dos pesos.

$$W \leftarrow W - \eta \cdot \nabla L \quad (2.9)$$

$$W^l \leftarrow W^l - \eta \cdot \frac{\partial L}{\partial W^l} \quad (2.10)$$

2.3 Análise Causal

A análise causal de *churn* envolve o estudo e a identificação das causas subjacentes à perda de clientes em um determinado negócio ou serviço. O objetivo é compreender os fatores que contribuem para o abandono de clientes, a fim de adotar medidas preventivas e estratégicas para reduzir a taxa de *churn*. Modelos econométricos, como diferenças em diferenças, variáveis instrumentais e até mesmo o design de regressão em descontinuidade, desempenham um papel crucial na identificação das causas do *churn*. Conforme destacado em Angrist e Pischke (2008), os modelos causais devem evidenciar as relações entre o que está sendo explicado e suas variáveis de controle. Dessa forma, o presente trabalho resume o conceito de efeitos de tratamento e os *frameworks* para encontrar tais valores.

2.3.1 Efeitos de tratamento

Nos estudos apresentados em Angrist e Pischke (2015) e Angrist e Pischke (2008), o efeito de tratamento refere-se à diferença observada em uma variável de resultado entre os grupos que receberam um tratamento e aqueles que não receberam. Esse conceito é central em estudos de pesquisa causal, onde os pesquisadores buscam entender o impacto causal de uma intervenção, política ou tratamento específico em uma determinada população. Pode-se resumir o efeito causal de uma intervenção como a variação na explicabilidade de uma variável em relação ao resultado apresentado no experimento realizado.

Além disso, a investigação considera a presença de heterogeneidade nos efeitos, reconhecendo que o impacto da intervenção pode variar entre diferentes subgrupos da po-

pulação. A análise aborda não apenas os efeitos imediatos, mas também os potenciais impactos de longo prazo da intervenção.

A validade interna e externa são cuidadosamente ponderadas para garantir conclusões confiáveis e relevantes no contexto específico do estudo, enquanto considerações éticas são integralmente incorporadas à condução da pesquisa, assegurando uma abordagem responsável e ética. Essas afirmações são consideradas padrões ouro na literatura e estão presentes em guias de experimentos, como o escrito por Duflo e Banerjee em Duflo e Banerjee (2017). Dessa forma, podemos classificar os efeitos de tratamento como efeito médio (ATE), efeito local (LATE), efeito condicional (CATE) e efeito heterogêneo (HTE).

Efeito de Tratamento Médio - ATE

A partir da literatura presente em Angrist e Pischke (2008), ATE representa a diferença média entre os resultados observados para um grupo de tratamento e um grupo de controle. Este indicador proporciona uma estimativa fundamental sobre o efeito causal médio da intervenção, permitindo-nos discernir se uma mudança observada na variável de resultado é atribuível à própria intervenção, fornecendo, assim, *insights* preciosos para a tomada de decisões e desenvolvimento de políticas. A estimação precisa do ATE requer a aplicação de métodos estatísticos rigorosos, sendo os experimentos controlados, nos quais os participantes são aleatoriamente designados para grupos de tratamento e controle, uma abordagem clássica.

Efeito de Tratamento Médio Local - LATE

O *Local Average Treatment Effect*, LATE, é um conceito fundamental na análise causal, especialmente em estudos que buscam entender os efeitos do tratamento em situações onde a adesão ao tratamento não é aleatória. Ao contrário do *Average Treatment Effect*, ATE, que reflete a média dos efeitos em toda a população, o LATE se concentra em subgrupos específicos de participantes que são "complacentes", ou seja, aqueles que rece-

beram o tratamento principalmente devido a características específicas. Essa abordagem permite uma compreensão mais refinada dos impactos do tratamento em situações onde a adesão é influenciada por fatores específicos, tornando-se particularmente relevante em contextos onde o tratamento pode afetar diferentes participantes de maneiras distintas.

Do conceituado pelos autores de Angrist e Pischke (2008), o LATE destaca-se por sua capacidade de capturar variações nos efeitos do tratamento, considerando a heterogeneidade na adesão e proporcionando uma visão mais precisa dos resultados causais em subgrupos específicos. Esta abordagem é valiosa em pesquisas que buscam discernir nuances nos efeitos das intervenções, reconhecendo que o impacto de um tratamento pode variar significativamente com base em características individuais ou contextuais. O LATE, portanto, oferece uma contribuição essencial para a compreensão das relações causais em estudos que enfrentam desafios de adesão não aleatória.

Efeito de Tratamento Condicional Local - CATE

O *Conditional Average Treatment Effect*, CATE, é um conceito chave na análise causal, direcionado para a compreensão dos efeitos do tratamento que variam em função de características específicas dos indivíduos ou condições de contexto, como definido pelos autores de Angrist e Pischke (2008) e Cunningham (2021). O CATE permite uma análise mais granular ao avaliar como o tratamento influencia diferentes subgrupos condicionalmente a certas variáveis.

Ao explorar o CATE, os pesquisadores podem investigar como diferentes características moderam os efeitos do tratamento. Por exemplo, em um estudo de intervenção de marketing para evitar *churn*, o CATE poderia revelar como a eficácia do tratamento varia entre clientes com diferentes gêneros e idades. Essa análise detalhada é crucial para personalizar estratégias de intervenção, identificando quais subgrupos podem se beneficiar mais ou menos do tratamento proposto.

Efeito de Tratamento Heterogêneo - HTE

Por meio do que é exemplificado em Molak (2023), a heterogeneidade nos efeitos de tratamento é um conceito crucial na análise causal, destacando que o impacto de uma intervenção pode variar substancialmente entre indivíduos ou grupos. Em vez de oferecer uma estimativa média do efeito de uma intervenção em toda a população, o *heterogeneous Treatment Effects*, HTE, direciona a atenção para a variação desses efeitos em subgrupos específicos. Isso reconhece a diversidade entre os participantes e sugere que o sucesso ou fracasso de uma intervenção pode depender de características individuais, contextuais ou de outros fatores moderadores.

A aplicação dos HTE na prevenção de *churn* emerge como uma estratégia inovadora e perspicaz no campo da gestão de clientes. Ao compreender que as respostas dos clientes às ações de retenção podem variar substancialmente com base em suas características individuais, o uso de HTE permite uma abordagem mais personalizada e eficaz. A análise minuciosa dessas características proporciona *insights* valiosos sobre os segmentos de clientes que são mais suscetíveis ou resistentes ao *churn*, permitindo a adaptação de estratégias específicas para atender às necessidades distintas de cada grupo.

Em última análise, o uso de HTE na prevenção de *churn* não apenas representa um avanço na gestão estratégica de clientes, mas também oferece uma vantagem competitiva ao reconhecer e atender de maneira precisa e eficiente às necessidades individuais dos clientes. Essa abordagem não apenas reduz a probabilidade de *churn*, mas também estabelece as bases para uma fidelização mais significativa, promovendo o crescimento sustentável das organizações em ambientes de negócios dinâmicos.

2.3.2 Meta-aprendizado

O emprego de meta-aprendizado, *metalearns*, em estratégias de análise causal de *churn* representa uma abordagem inovadora e sofisticada para entender os fatores sub-

jaçentes ao abandono de clientes. No contexto da previsão de *churn*, o meta-aprendizado transcende as técnicas tradicionais, capacitando os modelos a aprenderem como aprender com dados variados, adaptando-se dinamicamente a diferentes cenários e características dos clientes.

Em primeiro lugar, os *metalearns* oferecem a capacidade de otimizar a escolha e a configuração de modelos de aprendizado de máquina para prever o *churn* com base em características específicas dos clientes. Ao incorporar essa flexibilidade adaptativa, os modelos *metalearns* podem ajustar seus próprios hiperparâmetros e arquiteturas para otimizar a precisão da previsão em diferentes contextos com expresso em Molak (2023). Além disso, o meta-aprendizado permite uma melhor generalização para novos clientes ou mudanças nas condições de mercado. Os modelos podem aprender padrões gerais de comportamento de *churn*, tornando-os mais eficazes na previsão do abandono, mesmo para clientes cujas características não foram amplamente observadas durante o treinamento.

Modelos como a floresta causal apresentados em Wager e Athey (2018), os *metalearns* presentes em Chernozhukov et al. (2018) e Künzel et al. (2019), aliados ao conceito de *DeepIV* apresentado em Athey e Imbens (2019), são exemplos da utilização de conceitos de aprendizado de máquina na inferência causal.

X-learn

Definido conceitualmente em Künzel et al. (2019), pode-se resumir o *X-Learn* como uma metodologia avançada na análise causal, projetada para lidar com a complexidade inerente à inferência causal em configurações realistas. A essência do *X-Learn* está encapsulada em seu modelo duplo, que consiste em dois estágios principais. A partir do conceito apresentado por Künzel, no primeiro estágio, o modelo inicial é treinado para estimar a função de propensão ao tratamento, ou seja, a probabilidade de um indivíduo receber o tratamento dado seu conjunto de características. A equação fundamental para

este estágio é dada pela Equação 2.11.

$$P(T = 1|X) \tag{2.11}$$

No segundo estágio, o modelo utiliza o resultado observado e o tratamento real para calcular a "pseudo-outcome" e o "pseudo-tratamento" para cada indivíduo. Esses são construídos para simular o que teria acontecido se o indivíduo tivesse recebido o tratamento oposto ao que realmente recebeu. A pseudo-outcome é representada por Y_l e o pseudo-tratamento por T^* , onde $T^* = 1 - T$. Com a criação das pseudo-variáveis, um segundo modelo é treinado para prever o resultado com base na pseudo-outcome e no pseudo-tratamento. A equação resultante é evidenciada na Equação 2.12, na qual o modelo final é adaptado para capturar as nuances do tratamento e proporcionar uma estimativa robusta do efeito causal.

$$E(Y|T^*, X) \tag{2.12}$$

O *X-Learn* incorpora um conceito crucial chamado *balanced risk*, que busca minimizar a disparidade entre as estimativas de efeitos causais nos diferentes subgrupos da população como evidenciado por Facure (2020). A metodologia *X-Learn* frequentemente emprega técnicas de validação cruzada para avaliar seu desempenho, dividindo repetidamente os dados em conjuntos de treinamento e teste. O resumo de todo o modelo é visualizado na Figura 2.5.

Debiased/Orthogonal Machine Learning - DML

A metodologia *Debiased/Orthogonal Machine Learning*, DML, constitui-se de uma abordagem avançada na análise causal, visando superar desafios inerentes à inferência



Figura 2.5: Arquitetura do modelo de X-Learn. Fonte: Fature (2020).

causal em ambientes complexos. No cerne do DML está a busca por estimativas não viesadas dos efeitos causais, conforme definido no estudo de Chernozhukov et al. (2018), de forma a unir a praticidade e assertividade de modelos de aprendizado de máquina em previsões e a robustez ao erro de modelagens econométricas.

A característica fundamental do DML/OML é a introdução de um processo de desenvolvimento ortogonal, onde as variáveis de tratamento e resultado são desassociadas, reduzindo assim o viés potencial nas estimativas de efeito causal. O conceito de Frisch-Waugh-Lovell é central nesse processo, conforme expresso em Semenova et al. (2017) e Chernozhukov et al. (2018). É possível definir que o modelo linear é descrito como visto na Equação 2.13.

$$Y_{jt} = \beta_0 + T\theta + \beta_1 X + u \quad E[u|x, z] = 0 \quad (2.13)$$

Dessa forma, a Equação 2.13 pode ser ortogonalizada em relação ao erro e evidenciar estimadores consistente em raiz-n e não-viesados. Para tal, a Equação 2.13 é dividida em

três partes:

1. Regredir de Y em $X\beta_1X$, guardando os resíduos \widehat{Y}_{jt} ;
2. regredir de $T\theta$ em $X\beta_1X$, guardando os resíduos \widehat{T}
3. Regredir os resíduos dos pontos anteriores para encontrar o estimador $\widetilde{Y} = \theta\widetilde{T} + \varepsilon$

Assim, A fórmula geral do estimador DML pode ser expressa como:

$$\hat{\theta}_{DML} = \left(\frac{1}{n} \sum_{i=1}^n \frac{T_i - \hat{e}(X_i)}{\hat{e}(X_i)(1 - \hat{e}(X_i))} \cdot (Y_i - \hat{E}(Y_i|X_i)) \right) + \hat{\theta}_{OLS} \quad (2.14)$$

No qual os itens são expressos como:

- $\hat{\theta}_{DML}$: Estimativa DML do efeito causal.
- $\frac{T_i - \hat{e}(X_i)}{\hat{e}(X_i)(1 - \hat{e}(X_i))}$: Pesos resultantes do inverso da pontuação do modelo de propensão ao tratamento.
- Y_i : Valor observado da variável de resultado para a unidade i .
- $\hat{E}(Y_i|X_i)$: Estimativa do valor esperado da variável de resultado condicional ao conjunto de co-variáveis X_i obtido a partir do modelo de resultado condicional.
- $\hat{e}(X_i)$: Estimativa da probabilidade de tratamento condicional ao conjunto de co-variáveis X_i obtida a partir do modelo de propensão ao tratamento.
- $\hat{\theta}_{OLS}$: Estimativa do efeito causal obtida por meio de um modelo de regressão linear usando as variáveis de tratamento e co-variáveis como preditores.

A característica fundamental do DML é a introdução de um processo de desenvolvimento ortogonal, onde as variáveis de tratamento e resultado são desassociadas, reduzindo assim o viés potencial nas estimativas de efeito causal. Focando no cálculo do CATE como um indicador central do efeito causal, o *framework* permite avaliar como o tratamento impacta o resultado, levando em consideração a heterogeneidade nos dados.

Assim como o *X-Learn*, a modelagem frequentemente recorre a técnicas de validação cruzada para avaliação de desempenho, garantindo que o modelo seja capaz de generalizar

para novos dados. Dessa forma, o modelo pode garantir que sua análise causal é robusta ao erro e não apresenta viés em nenhuma variável de controle. Por conseguinte, podemos expressar tal modelagem como visto na Figura 2.6, onde o processo é repetido K vezes para garantir a robustez e a ausência de viés no resultado.

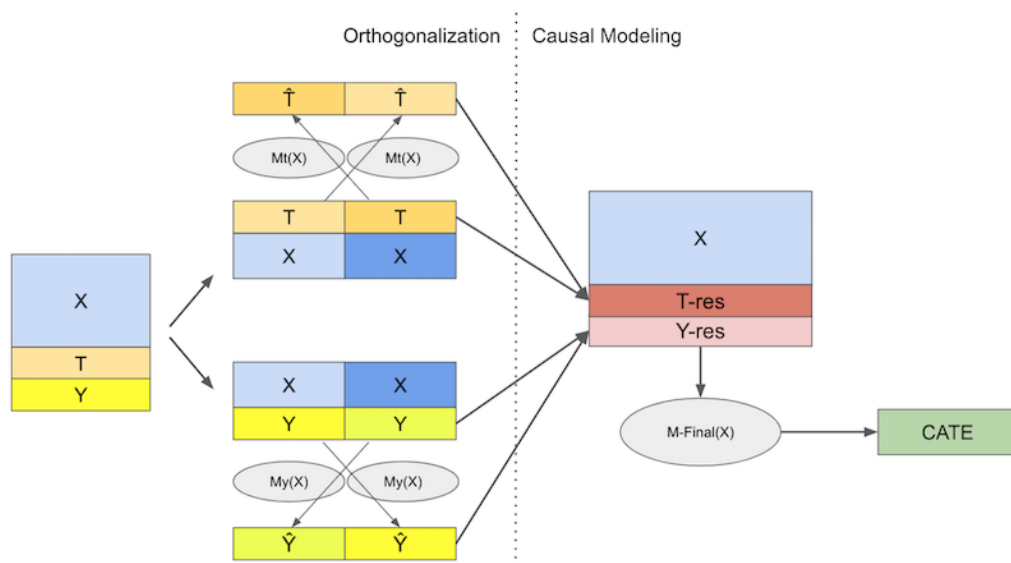


Figura 2.6: Arquitetura de modelo de DML. Fonte: Facure (2020).

2.4 Métricas Para Classificação

A avaliação de modelos de classificação em aprendizado de máquina é crucial para compreender sua eficácia na atribuição de rótulos às instâncias de dados. Métricas como acurácia, precisão, sensibilidade, score F1 e área sob a curva ROC são comumente utilizadas para quantificar diferentes aspectos do desempenho do modelo e sua validação como proposto por Schneider (2016). A acurácia reflete a proporção de predições corretas, enquanto a precisão enfoca a minimização de falsos positivos e a sensibilidade, a identificação de falsos negativos. O score F1 combina precisão e sensibilidade em uma única métrica de média harmônica, e a área sob a curva, AUC, representa a capacidade de distinguir entre classes. Além disso, em contextos de inferências causais, métricas como *uplift*, ganhos acumulados e estatística Kolmogorov-Smirnov (KS) são empregadas para avaliar o desempenho do classificador, considerando impacto positivo, eficácia em relação a uma abordagem aleatória e capacidade de separação entre distribuições, respectivamente. Ao avaliar modelos, é essencial escolher métricas alinhadas aos objetivos específicos do problema e considerar o contexto da aplicação.

Acurácia

A acurácia avalia a precisão global do modelo ao prever corretamente as classes das instâncias de dados. A fórmula da acurácia é representada pela Equação 2.15 exemplificada em Schneider (2016). Expressa como uma proporção ou porcentagem, fornece uma visão geral da capacidade do modelo em acertar as previsões. Embora seja uma métrica intuitiva e fácil de interpretar, é crucial utilizá-la com cautela em conjuntos de dados desbalanceados, onde a distribuição de classes não é uniforme. Em tais casos, a acurácia pode ser enganosa, destacando a importância de considerar métricas adicionais, como precisão, escore F1 e a área sob a curva, para uma avaliação mais abrangente do desempenho do modelo.

$$\text{Acurácia} = \frac{\text{Número de Predições Corretas}}{\text{Número Total de Predições}} \quad (2.15)$$

Precisão, Sensibilidade e Score F1

A precisão é uma métrica crucial em problemas de classificação, concentrando-se na qualidade das predições positivas feitas por um modelo, como representa a Equação 2.16. Essa métrica é particularmente relevante quando o foco está na minimização de falsos positivos, ou seja, quando é crucial garantir que as instâncias classificadas como positivas pelo modelo realmente pertençam a essa categoria. A precisão é expressa como uma proporção ou porcentagem, proporcionando uma visão clara da proporção de predições positivas corretas em relação ao total de predições positivas. Os conceitos para cada métrica partem do que é definido e expresso pelo autor de Geron (2019).

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Positivos}} \quad (2.16)$$

A sensibilidade, ou *recall*, é a métrica que concentra na capacidade de um modelo identificar corretamente todas as instâncias positivas no conjunto de dados. Essa métrica é particularmente valiosa em situações em que a omissão de instâncias positivas é de grande importância, como em problemas médicos, onde a não detecção de casos positivos pode ter consequências graves.

A fórmula para a sensibilidade, representada pela Equação 2.17, destaca a proporção de instâncias positivas reais que o modelo foi capaz de identificar em relação ao total de instâncias positivas existentes. Em termos práticos, a sensibilidade fornece informações sobre a eficácia do modelo em capturar casos relevantes, evitando falsos negativos. É importante mencionar que, ao otimizar um modelo para aumentar a sensibilidade, geralmente há uma diminuição correspondente na precisão. Isso ocorre porque o modelo tende a classificar mais instâncias como positivas para garantir que não perca casos verdadeiros.

$$\text{Sensibilidade} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Negativos}} \quad (2.17)$$

O escore F1 é uma métrica que combina precisão e sensibilidade, oferecendo a média harmônica entre a capacidade de um modelo de classificação em realizar previsões corretas e identificar todas as instâncias positivas. Essa métrica é particularmente útil quando há a necessidade de equilibrar precisão e sensibilidade, sendo valiosa em cenários onde ambas são importantes para a avaliação do desempenho do modelo.

A fórmula do escore F1, representada pela Equação 2.18, destaca a média harmônica como uma abordagem que penaliza mais fortemente valores extremos. Isso significa que o escore F1 é mais sensível a desequilíbrios entre precisão e sensibilidade do que a média aritmética, fornecendo uma medida mais equilibrada do desempenho do modelo em comparação com métricas que enfatizam apenas uma das duas. Ao utilizar o "*F1-Score*", os tomadores de decisão podem tomar decisões mais informadas sobre o desempenho global de um modelo, considerando tanto a capacidade de realizar previsões corretas quanto a capacidade de identificar todas as instâncias positivas, oferecendo uma avaliação mais completa e equilibrada.

$$\text{Escore-F1} = \frac{2 \times (\text{Precisão} \times \text{Sensibilidade})}{\text{Precisão} + \text{Sensibilidade}} \quad (2.18)$$

Curva Sobre Área - AUC

A área sob a curva (AUC) representa graficamente a taxa de verdadeiros positivos em relação à taxa de falsos positivos para diferentes limiares de decisão do modelo, quantificando numericamente a qualidade do modelo em classificar instâncias positivas e negativas. Essa métrica fornece uma medida global da capacidade discriminativa do modelo, sendo especialmente útil em problemas de classificação binária. A fórmula para calcular a AUC, representada pela Equação 2.19, demonstrada pelo autor de Geron (2019),

destaca a integral da curva ROC (*Receiver Operating Characteristic*), que é um gráfico que ilustra a taxa de verdadeiros positivos em função da taxa de falsos positivos. Quanto maior a AUC, melhor o desempenho do modelo em distinguir entre as classes. A AUC é uma métrica robusta que não depende de um limiar de decisão específico, sendo útil em situações em que a escolha do ponto de corte para predições probabilísticas é desafiadora.

$$\text{AUC-ROC} = \int_0^1 \text{Sensibilidade}(\text{Especificidade}) d\text{Especificidade} \quad (2.19)$$

Curva de Levantamento - *Lift Plot*

O *Lift Plot* é uma ferramenta gráfica amplamente utilizada na análise de modelos de classificação, especialmente em contextos de marketing e retenção de clientes. Sua utilidade reside na avaliação do desempenho de um modelo em comparação com uma abordagem aleatória ou de referência. O gráfico compara a taxa de sucesso acumulada do modelo em relação à taxa de sucesso acumulada esperada caso estivéssemos fazendo previsões aleatórias. Os autores de Jaffery e Liu (2009) resumem o conceito ao conhecimento de quais clientes estão mais suscetíveis a políticas de novos bens ou serviços.

No *Lift Plot*, o eixo x geralmente representa a fração acumulada dos dados, como os top 10% dos casos classificados pelo modelo, enquanto o eixo y representa o *lift*. O levantamento é a razão entre a taxa de sucesso do modelo e a taxa de sucesso aleatória esperada. O valor do *lift* em um ponto específico indica quantas vezes o modelo supera o desempenho aleatório nessa fração dos dados como definido pelos autores Jaffery e Liu (2009). A figura 2.7 demonstra tal relação.

Essa visualização é particularmente valiosa em situações onde o objetivo é melhorar a eficácia de estratégias de marketing, permitindo uma compreensão clara de como o modelo contribui para o aumento da taxa de sucesso em comparação com uma abordagem aleatória. O *Lift Plot* facilita a identificação das áreas em que o modelo se destaca, proporcionando insights importantes para otimizar estratégias e alocar recursos de maneira

mais eficiente.

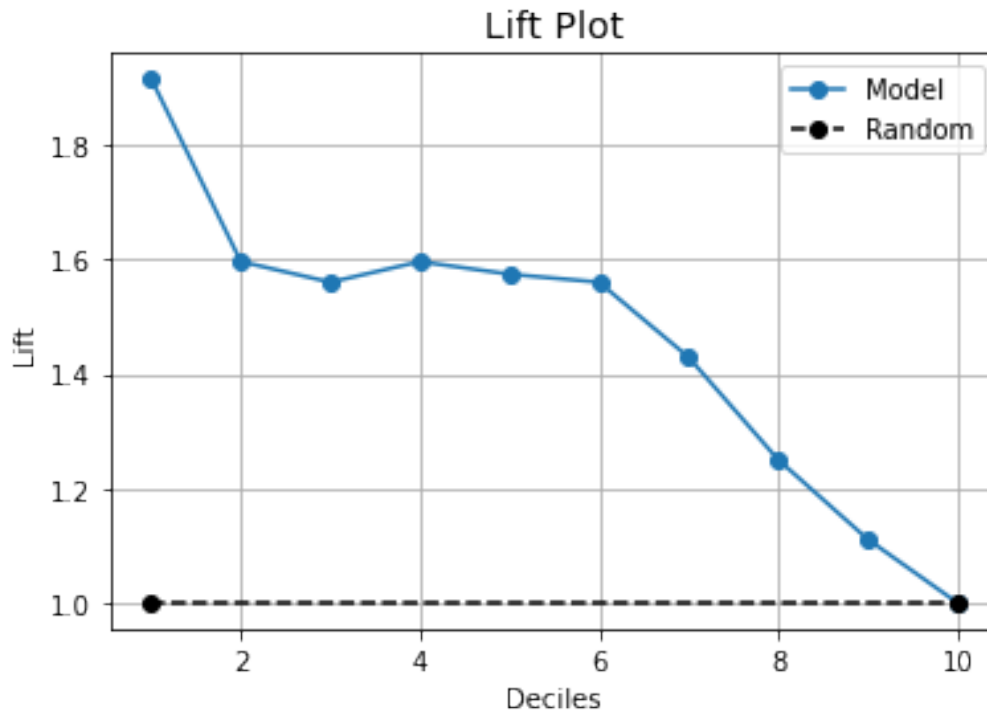


Figura 2.7: Gráfico de *lift* de um modelo de aprendizado de máquina. Fonte: Jaffery e Liu (2009).

Ganho Cumulativo - *Cumulative Gain*

O Ganho Cumulativo é uma métrica usada para avaliar o desempenho de modelos de classificação binária ou sistemas de recomendação, especialmente em cenários onde o foco está na identificação de instâncias positivas. Essa métrica é particularmente útil em problemas de classificação desbalanceada, nos quais as classes positivas são significativamente menos frequentes.

O objetivo do Ganho Cumulativo, exemplificado pelos autores de Jaffery e Liu (2009), é avaliar o aumento acumulado à medida que você percorre a lista classificada. Em resumo, quanto mais rapidamente você encontra as instâncias positivas, maior é o ganho cumulativo. A Equação 2.20 demonstra o somatório necessário para a determinação do

ganho, representando a diferença acumulada entre o modelo e uma abordagem aleatória, à medida que mais instâncias são consideradas na lista classificada.

$$\text{Ganho Cumulativo} = \sum_{i=1}^n \text{Ganho}(i) \quad (2.20)$$

Essa métrica proporciona uma visão clara de como o modelo se destaca em identificar instâncias positivas em comparação com uma abordagem aleatória, facilitando a interpretação do desempenho em cenários onde a classe positiva é rara. A demonstração desse conceito é vista na Figura 2.8 que expõem o modelo aleatório, o modelo com sobre ajuste e o modelo testado.

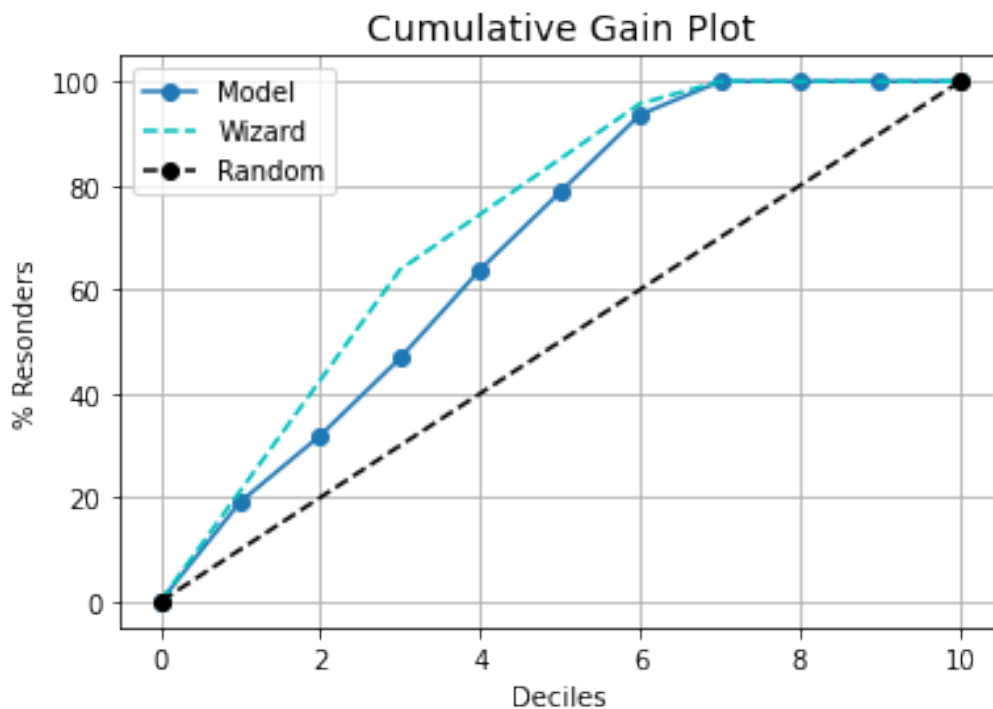


Figura 2.8: Gráfico de ganho cumulativo para um modelo de floresta aleatória. Fonte: Jaffery e Liu (2009).

Estatística Kolmogorov-Smirnov - KS

A conceituação inicial dos autores de Dimitrova et al. (2020) exemplifica que a Estatística Kolmogorov-Smirnov (KS) é uma métrica estatística amplamente empregada para avaliar a capacidade discriminativa de um modelo de classificação binária, análise de diferença entre duas distribuições. Essa métrica é especialmente útil em problemas caracterizados por desequilíbrio nas classes, como detecção de fraude, identificação de anomalias ou avaliação de risco de crédito. A fórmula da KS, exposta na Equação 2.21, destaca o cálculo da diferença máxima entre as taxas de falso positivo ($FPR(c)$) e verdadeiro positivo ($TPR(c)$) em diferentes pontos de corte na curva ROC.

$$KS = \max|FRP(c) - TPR(c)| \quad (2.21)$$

A KS é derivada da curva ROC e é definida como a maior diferença vertical entre as curvas ROC para as classes positiva e negativa. Essa métrica oferece uma vantagem significativa, pois fornece uma única medida que resume a eficácia global do modelo em separar as duas classes. Um valor mais alto de KS indica um melhor desempenho do modelo em distinguir entre instâncias positivas e negativas. Exemplo do uso de KS é apresentado na Figura 2.9.

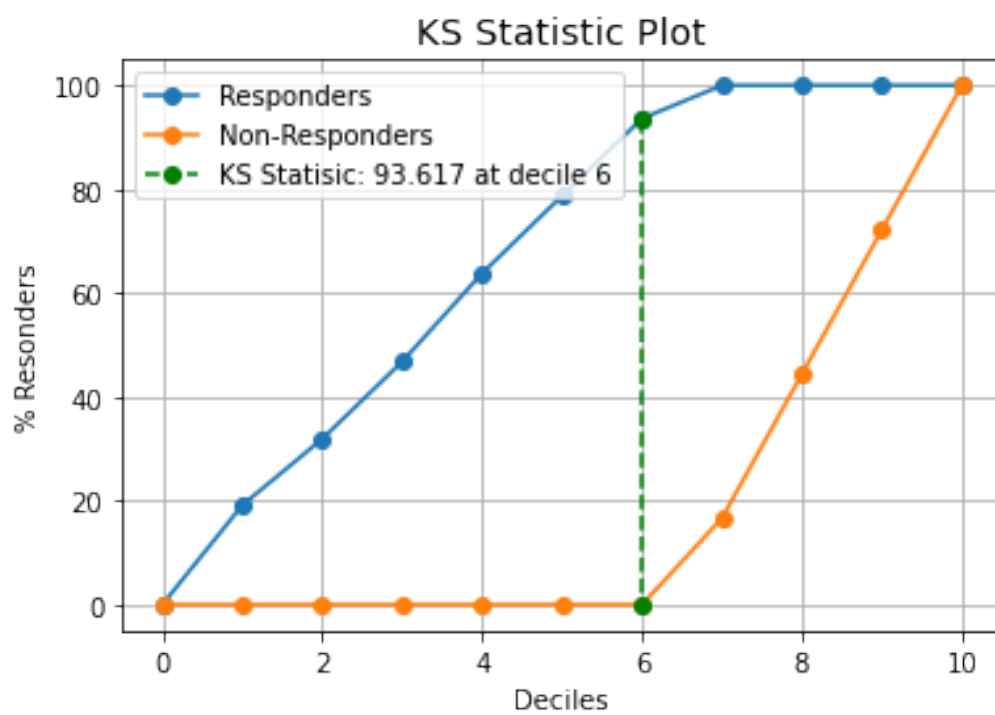


Figura 2.9: Gráfico de ganho cumulativo para um modelo de floresta aleatória. Fonte: Jaffery e Liu (2009).

Capítulo 3

Trabalhos Relacionados

3.1 Literatura do Ciclo de Vida do Cliente

Há vasta literatura de CRM apresentada em Kurtz e Clow (1998), os conceitos de *churn*, sua previsão e seu entendimento sedimentados por Wu et al. (2021) e *DEVRI-ENDT2021497* o presente trabalho pretende expandir a literatura em conceitos de previsão e explicação do *churn*. Por conseguinte, nesta seção trabalhos relacionados a tais objetivos são apresentados e destrinchados para o entendimento de sua modelagem. No sentido estreito, todos os estudos que fornecem base a este texto tem como motor o uso de aprendizado de máquina seja para previsão, por meio de modelagem de classificação, ou seja por análise causal de implementação de políticas de marketing.

A construção do arcabouço teórico parte dos trabalho de revisão sistemática realizado em Schneider (2016), do trabalho Veloso (2012) realizado em industrias de alto giro e no uso da mesma base de dados apresentada. Logo, a seção é dividida entre literatura de língua inglesa e de língua portuguesa. Ademais, conceitos de previsão e de análise causal estão separados em seções diferentes devido a interligação dos trabalhos de cada área.

3.1.1 Previsão de Churn

No contexto de previsão de *churn* o trabalho realizado pelos autores de Chiang et al. (2003) propõem um metodologia "*goal oriented*" sequencial para a análise e previsão

do *churn* de forma a apresentar resultados que sejam palpáveis a tomadores de decisão no mercado estudado. Utilizando de conceito de metas e associação o trabalho Chiang et al. (2003) pode definir metas alvos e buscar compreender a perda de clientes em ambientes financeiros demonstrando. Contudo, o modelo é pouco reproduzível e apresenta viés de aplicado a outros tipos de industria.

De modo a expandir o conceito anterior, o estudo Hung et al. (2006) apresenta um novo *framework* para a previsão e tratamento da propensão de *churn* de clientes de serviços de telecomunicações em um mercado com menos regulações, o estudo verifica dados de Taiwan. Os autores usam métodos de árvores de decisão e redes neurais artificiais para encontrar possíveis ações atratoras para a perda de clientes. Os resultados apresentam números mais precisos de quais clientes estão saindo e como cada variável interfere nesse processo.

De forma contrária, os autores de Neslin et al. (2006) demonstram que, antes da utilização de modelos matemáticos e estatísticos robusto, o cerne metodológico do que pode causar o *churn* é importante. O estudo utilizou dados da acurácia de modelos de previsão de *churn* para demonstrar que modelos com melhor conceituação teoria possuem métricas mais aprimoradas. Neslin et al. (2006) dialoga que pesquisas de caráter quantitativo devem utilizar de conceitos robustos qualitativos.

De forma a expandir o campo de métodos para previsão de *churn* para base de dados desbalanceadas, os autores de Xie et al. (2009) apresentam uma modelagem baseada em florestas aleatórias balanceadas, utilizando modelos de balanceamento de base como o "*oversampling*". O estudo demonstra que a utilização de tais florestas apresentam melhores métricas de acurácia do que outras modelagens como redes neurais e árvores de decisão no contexto de rebalanceamento dos dados de saída de clientes. O estudo também dialoga que a mudança pode trazer diferenças na inferência final.

De forma continuada, no campo de modelos robustos a desbalanceamento de dados, os autores de J. Burez (2009) corroboram com Xie et al. (2009) ao demonstrar que a

utilização de técnicas de balanceamento produzem modelos com melhores resultados em métricas de acurácia. No estudo os autores utilizam do conceitos de "*undersampling*" e de modelos de florestas aleatórias "*weighted*" e "*gradient boosting*", cujo oferecem acurácias mais altas e demonstram curvas AUC que concordam com essa melhora. Assim, apresentado modelos que possam ser utilizados por tomadores de decisão.

Os autores de Huang et al. (2012) apresentam estudo sobre o *churn* no contexto de telecomunicações utilizando os conceitos de rebalanceamento citados. No tocante ao inédito, o estudo demonstra que um novo vetor de variáveis, que é composto desde custos de ligação e métodos de pagamento, desenvolve melhores resultados que o modelos utilizado anteriormente pela companhia, dados de perfil do cliente e dados de ligações agregados. Ademais, os autores citam que modelos de "*support vector machine*" e árvores de decisão tem melhores métricas de acurácia.

No contexto de aprendizado profundo, o estudo Castanedo et al. (2014) utiliza de bilhões de observações de dados de ligações para prever *churn* de clientes de planos pré pagos. O autor pode encontrar uma acurácia de 80% com seu modelo, de mostrando sua robustez. Seguindo nesta área, os autores de Mishra e Reddy (2017) expandem o conceito anterior ao utilizarem arquiteturas de redes neurais convolucionais a bases semelhantes. O modelo de Mishra e Reddy (2017) demonstra apresenta alta acurácia, 86%, e alto score F1, 92%. O estudo Agrawal et al. (2018) demonstra que além de métricas a modelagem em "*deep learning*" deve demonstrar explicabilidade. Dessa forma, há melhoria no processo de tomada de decisão pelo meio gerencial.

Os autores de Wu et al. (2021) aferem que clientes propensos a abandonar o produto podem ser agrupados em conglomerados facilitando a previsão dos mesmos. Para tal os autores utilizam de três base de dados diferentes.

De início, a primeira é baseada na base sintética "*TELCO customer churn*" da IBM, mesmo base utilizada pela presente investigação, onde dados categóricos de perfil de cliente e contratos são alinhados a dados de consumos contínuos. A segunda base

relaciona dados geoespaciais e estatísticas de tempo e quantidade de ligações. Por fim, a terceira base é baseada em dados de perfis de consumidor e dos dados presentes na segunda base.

A pesquisa realizada em Wu et al. (2021) utiliza de conceitos de "*oversample*" para o balanceamento dos indicadores de *churn*. Contudo, os resultados são apresentados com e sem o uso de tal ferramenta para demonstrar a calibração de probabilidade de cada modelo para cada base.

Dos resultados referentes a primeira base de dados, os resultados utilizando os modelos de regressão logística e ADABOOST obtiveram melhores métricas de AUC-ROC e acurácia. Porém, o estudo Wu et al. (2021) alerta que há baixa razão de precisão-sensibilidade podendo seguir na presença de falsos positivos invalidando o uso do modelo.

Literatura Lusófona

No contexto Lusófono, pode-se citar o trabalho dos autores de Botelho e Tostes (2010), os quais apresentam modelo de regressão logística para a previsão de *churn* em uma administradora de cartões de crédito no Brasil. Posteriormente, há a apresentação de quais variáveis explicativas demonstram a maior probabilidade de um consumidor sair da plataforma.

De forma continuada, a dissertação de Veloso (2012), demonstra a utilização de metodologias de aprendizado de máquina para a previsão de saída de clientes em uma indústria de retalhos. Veloso (2012) define que modelos de florestas aleatórias possuem melhores métricas para o contexto estudado.

De forma expandida, o estudo de Franceschi (2019) aplica as modelagens propostas por Veloso (2012) ao contexto bancário no Brasil e demonstra que bases de dados desbalanceadas podem apresentar métricas mais baixas; porém, robustas a inferência de tomadores de decisão.

Adicionalmente, Araújo (2022) estudo o conceito de *churn* em clientes de seguro de

vida do banco do Brasil no contexto da análise de sobrevivência. Dessa forma, José Araújo exemplifica que a utilização de regressão de Cox aliado a o estimador Kaplan-Meier trazem novos insights ao campo.

3.1.2 Modelos de elevação - Análise Causal

Aliado ao conceito de ciclo de vida do usuário de um produto, tomadores de decisão seguem caminhos para a manutenção da sua base de cliente de forma a tornar o ecossistema da solução mais sustentável tecnicamente e financeiramente como exposto pelos autores de Kurtz e Clow (1998). Dessa maneira, o campo de análise causal de políticas em varejo apresenta soluções ao problema de *churn*.

Seguindo o campo puramente de análise causal, em Verhelst et al. (2020) os autores apresentam um caminho de pesquisa que alia o conceito de inferência causal com a utilização de aprendizado de máquina para definições das causas de uma desistência em clientes de telecomunicações. Primeiro, os autores utilizam-se de florestas aleatórias para criação de um modelo preditivo. Segundo, definem conceitos de análise causal para à avaliação das causas de churn na base analisada (Custos Adicionais e Planos Não Adequados). Por fim, realizam teste de sensibilidade em suas variáveis de controle para definir causas não monótonas da perda de clientes. Adicionalmente, os estudos de Verhelst et al. (2021a) expandem o contexto de Verhelst et al. (2020) ao combinar modelos de análise causal a modelos de previsão de churn.

O estudo presente em Verhelst et al. (2021a) expande os conceitos expostos no artigo citado anteriormente. Os autores propõe a utilização de modelos de elevação ("*Uplifting Models*") para tentar evitar o churn de uma companhia de telecom. Neste contexto, o estudo define um experimento a partir das criação de um ranqueamento de possíveis evasões da plataforma e uma política de contato com tais clientes para a resolução de problemas encontrados. O autores, contudo, expõem que não houve mudança significativa de métricas entre o uso do modelo preditivo unicamente e o uso do modelo de elevação.

Os autores de Devriendt et al. (2021) defendem que a utilização de "*Uplifting Models*" pode ser melhor que o uso de modelos preditivos de churn. Para tal, o estudo define métrica que avalia o aumento de lucro devido a criação de políticas de intervenção em grupos de tratamento de clientes. Assim, os autores conseguem exemplificar que, possivelmente, experimentos bem definidos, cujo o grupo de tratamento é bem definido, podem incrementar mais o lucro do que unicamente persuadir clientes muito próximos a sair.

De forma adicional, Rudd et al. (2021) aborda a questão crítica do *churn* de clientes, que se refere à perda ou redução do envolvimento dos clientes com uma empresa ao longo do tempo. O estudo propõe um modelo de propensão ao *churn*, combinando uma rede neural profunda com uma análise causal bayesiana, visando oferecer uma abordagem inovadora para enfrentar esse desafio.

A singularidade desse trabalho reside na aplicação dessa abordagem em larga escala a conjuntos massivos de dados financeiros, com foco particular em empresas de previdência. Os resultados dos experimentos conduzidos em 12 conjuntos de dados validam a eficácia do modelo proposto, evidenciando seu desempenho superior em comparação com sete classificadores padrão, a partir de uma firma financeira local. Além da validação do modelo, a análise de causalidade destaca fatores críticos que influenciam o *churn* de clientes, incluindo o saldo da conta, o crescimento da conta e o gênero. Em suma, o estudo conclui ressaltando a importância de gerenciar proativamente o *churn* da partir de variáveis atratoras.

Capítulo 4

Modelagem

O mundo tecnológico passa por uma rápida evolução e apresenta disrupções circulares, sendo o aprendizado de máquina um claro exemplo desse fenômeno. Nesse contexto, o presente trabalho adota uma abordagem fundamentada nos princípios de desenvolvimento ágil, assim como na manutenção da qualidade, conforme explorado em Studer et al. (2020). No entanto, é imperativo integrar as ferramentas ao conceito teórico estudado. Por conseguinte, expandindo o que é estudado em Wu et al. (2021) e Rudd et al. (2021) no estudo de previsão e causa do *churn*. Dessa forma, a pesquisa é centrada no desenvolvimento da ferramenta de previsão e análise de causa do *churn* aliando modelo de *machine learning* e análise causal. Para tal, duas proposições são levantadas:

- Proposição 1: Características do perfil do cliente, como idade, gênero ou estado civil, exercem maior influência no *churn* do que a forma de pagamento ou o tipo de contrato;
- Proposição 2: A oferta de contratos duradouros impacta na redução do *churn*.

Assim, a criação da modelagem será estruturada de acordo com os princípios do CRISP-ML(Q), definindo requisitos e limitações, identificando riscos e propondo métodos para superá-los em áreas-chave. Estas áreas compreendem o conhecimento dos dados e do domínio de negócios, a manipulação de dados, a engenharia do modelo de aprendizado de máquina, a verificação da qualidade do modelo, bem como a entrega e manutenção.

4.1 Conhecimento de Dados, Negócio e Manipulação de dados

O presente trabalho usará a base de dados "TELCO customer churn 11.1.3" com cerca de 7043 observações para 50+ variáveis, fornecida pela IBM. A base conta com dados sintéticos de uma empresa de telecomunicações fornecendo desde dados geo-referenciados dos clientes, passando por dados de perfis de contratos e usuário, até indicadores de valoração do usuário (*CLTV*, *churn*, causa do *churn*).

4.1.1 Dados e Negócio

O entendimento do negócio desempenha um papel fundamental na organização dos dados e, subsequentemente, na modelagem para compreensão do *churn*. Portanto, é crucial reconhecer que a base de dados integral é composta por cinco grandes áreas: demografia, localização, população, serviços e status. Para este estudo, foi criada uma nova base de dados que integra as áreas de população, serviços e status. As variáveis selecionadas podem ser classificadas como categóricas, conforme referenciado na tabela 4.1, e numéricas, conforme apresentado na tabela 4.2.

Variável	Descrição	Trabalhos relacionados
CustomerID	identificação do cliente único (não utilizado)	Hung et al. (2006), Huang et al. (2012), Wu et al. (2021), Veloso (2012)
ChurnLabel	Identifica se o contratante deixou a plataforma do produto	Utilizada em todos os trabalhos como marcador
Gender	Gênero do contratante (Feminino e Masculino)	Hung et al. (2006), Huang et al. (2012), Wu et al. (2021)
SeniorCitizen	Identifica se o contratante tem mais que 65 anos de idade	Hung et al. (2006), Huang et al. (2012), Wu et al. (2021)
Partner	Identifica se o contratante é casado	Wu et al. (2021), Veloso (2012)
Dependents	Identifica se o contratante possui dependentes (filhos, pais ou avós)	Wu et al. (2021), Veloso (2012)
PhoneService	Identifica se o contratante possui serviço de telefone fixo	Huang et al. (2012), Castanedo et al. (2014), Wu et al. (2021)
MultipleLines	Identifica se o contratante possui múltiplas linhas de telefone fixo	Huang et al. (2012), Wu et al. (2021)
InternetService	Identifica se o contratante possui serviço de internet (Fibra ótica, Cabo Coaxial, DSL ou Não)	Wu et al. (2021)
OnlineSecurity	Identifica se o contratante possui serviço de proteção de perigos cibernéticos	Wu et al. (2021)
OnlineBackup	Identifica se o contratante possui serviço de backup	Wu et al. (2021)
DeviceProtection	Identifica se o contratante possui serviço de equipamentos fornecidos pela empresa	Wu et al. (2021)
TechSupport	Identifica se o contratante possui serviço prioritário de suporte técnico	Wu et al. (2021)
StreamingTV	Identifica se o contratante utiliza serviço de streaming de TV	Wu et al. (2021)
StreamingMovies	Identifica se o contratante utiliza serviço de streaming de filmes	Wu et al. (2021)
Contract	Identifica o tipo de contrato (Mês-a-Mês, Um Ano ou Dois Anos)	Wu et al. (2021)
PaperlessBilling	Identifica se o contratante optou pelo uso ou não de papel para a conta	Wu et al. (2021)
PaymentMethod	Identifica o método de pagamento do contratante (Cheque, Débito em conta ou Cartão de crédito)	Huang et al. (2012), Wu et al. (2021)
Offer	Identifica se o contratante recebeu uma oferta de desconto	Variável criada

Tabela 4.1: Variáveis categóricas da base de dados

Variável	Descrição	Trabalhos relacionados
TenureMonths	Quantidade de meses que o contratante está utilizando os serviços	Hung et al. (2006), Huang et al. (2012), Wu et al. (2021), Verhelst et al. (2020)
TotalCharges	Valor total em Dólar das contas já pagas pelo contratante	Huang et al. (2012), Wu et al. (2021)
AvgMonthlyGBDownload	Média mensal de gigabytes utilizados em descargas	Verhelst et al. (2020)
MonthlyCharges	Valor da conta mensal	Hung et al. (2006), Huang et al. (2012), Wu et al. (2021)

Tabela 4.2: Variáveis numéricas da base de dados

Análise Exploratória dos dados Categóricos

Para uma modelagem mais eficaz, é essencial compreender os dados categóricos contidos na base de dados. Nesse sentido, os dados serão analisados de maneira a fundamentar a melhor solução de aprendizado de máquina para a identificação do *churn*.

Inicialmente, os dados relativos ao perfil do cliente são apresentados de maneira a agrupá-los de forma distinta. Quanto ao gênero, a variável é binária, contendo valores masculino e feminino com uma distribuição relativamente equilibrada, aproximadamente 50,5% sendo do gênero masculino. No que diz respeito à quantidade de clientes idosos, observa-se a predominância de clientes com idade inferior a 65 anos, representando cerca de 83,8% de toda a base, onde o valor 0 para essa variável é prevalente. Em seguida, a distribuição de clientes que possuem parceiro ou são casados assemelha-se à questão de gênero, onde clientes sem laços matrimoniais representam 51,7% do total da base. Por fim, a maioria dos clientes não possui dependentes, com 77% da base apresentando o valor 0 para essa variável. A distribuição dessas variáveis, segregadas entre *churn* ou não, está representada na figura 4.1.

Os serviços contratados também exibem uma distribuição equilibrada. Focando nos serviços de internet, a base de dados é majoritariamente composta pelo uso de fibra ótica, representando 44%, seguido pelo uso de DSL, com 34,4%. No que se refere aos serviços de *streaming*, aproximadamente 40% são utilizados para televisão, enquanto 38,8% são destinados a filmes. Além disso, a contratação do serviço prioritário de suporte é considerada minoritária, com cerca de 29% dos contratantes optando por essa opção. A distribuição, segregada entre *churn* ou não, pode ser visualizada na figura 4.2. É válido pontuar que clientes que usam fibra ótica tendem a apresentar maior probabilidade de

churn dado uma análise primária do gráfico.

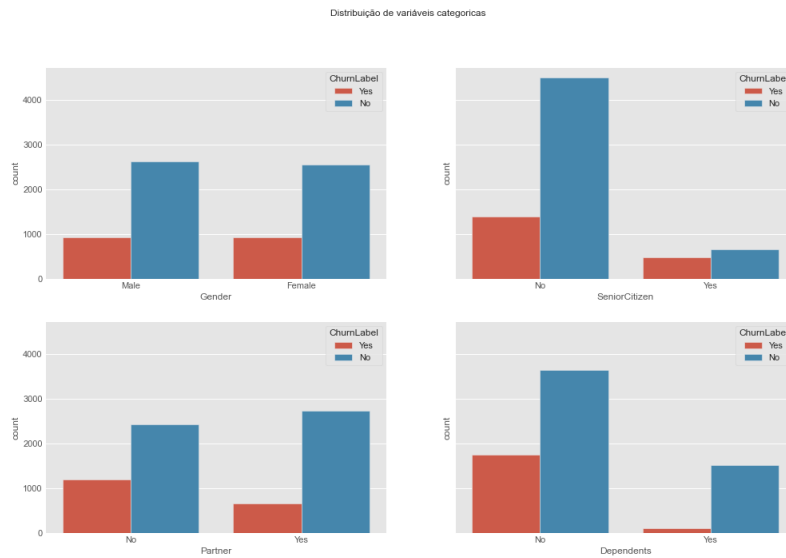


Figura 4.1: Distribuição de variáveis categóricas do perfil do cliente.

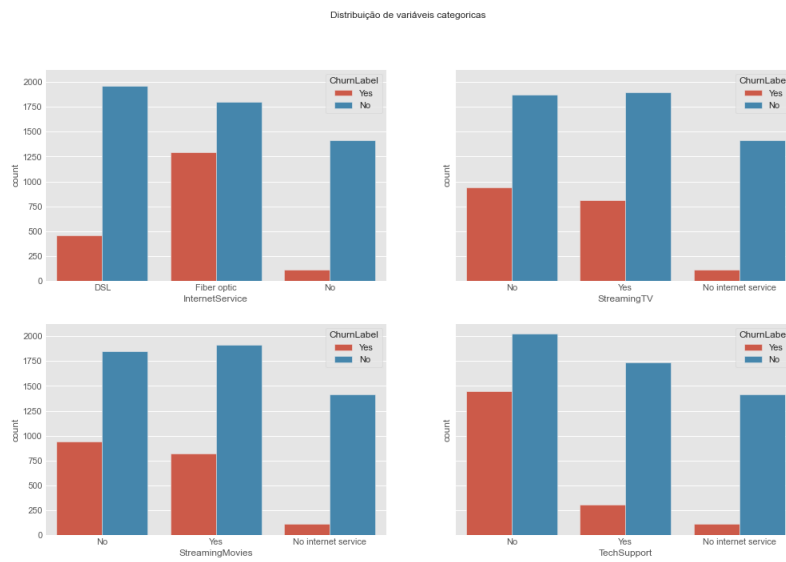


Figura 4.2: Distribuição de variáveis categóricas da utilização de serviços.

Dos aspectos contratuais, observa-se que a base é majoritariamente composta por contratos de mês-a-mês, abrangendo 55% do total da base, e por cheques eletrônicos, que representam 33,6% como a forma prioritária de pagamento. Cerca de 59% dos clientes

indicam preferir não receber suas cobranças em papel. Além disso, 55% dos clientes não possuem nenhuma oferta especial vinculada aos seus contratos. As distribuições, com estratificação para o *churn*, podem ser visualizadas na figura 4.3. Contratos do tipo mês-a-mês tendem a apresentar maior propensão ao *churn*, assim como o uso de cheques eletrônicos como forma de pagamento. De forma adicional, a oferta E aparenta não ser efetiva.

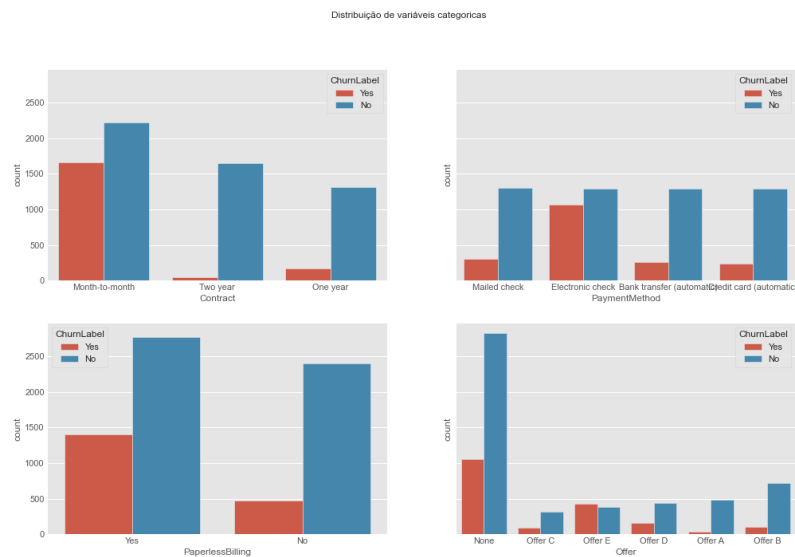


Figura 4.3: Distribuição de variáveis categóricas referentes a especificações contratuais.

De forma resumida, a maioria das variáveis categóricas possuem equidade em sua distribuição. Contudo, variáveis que definem características pertinentes do usuário, como idade, tipo de pagamento e tipo de contrato, apresentam desbalanço e demonstram predileção a deixar a plataforma.

Análise Exploratória dos dados Numéricos

Ao analisar as variáveis numéricas, percebe-se que o estudo do custo do serviço ao cliente, assim como a utilização total do serviço de internet, pode proporcionar insights mais aprofundados sobre as possíveis causas do churn.

Primeiro, a média da quantidade de meses de um contratante é de aproximadamente

33, com uma mediana de 29, evidenciando que os clientes tendem a permanecer por mais de dois anos utilizando os serviços. No entanto, ao realizar uma análise mais detalhada dos dados relacionados ao *churn*, observa-se que a média e a mediana do número de meses diminuem para 18 e 10, respectivamente. Isso sugere que os clientes que tendem a abandonar o produto o fazem de forma rápida, evidenciando uma possível correlação entre os tipos de contrato e a saída do cliente. A figura 4.4 explora a possível não normalidade da distribuição dessa variável, aliada à propensão do *churn* em relação à quantidade de meses que o cliente utiliza o serviço.

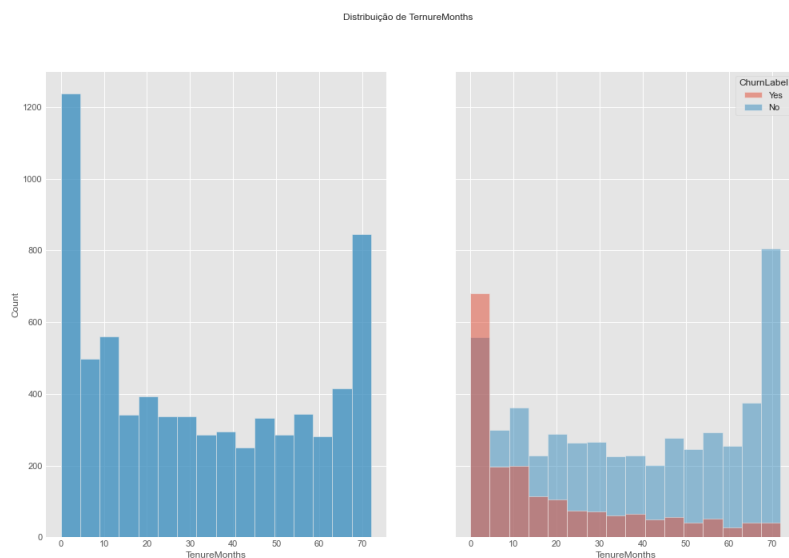


Figura 4.4: Distribuição da quantidade de meses dos clientes na plataforma e sua propensão a saída.

Segundo, o total monetário pago indica a permanência do contratante. Relacionada a quantidade de meses que cliente permanece na estrutura do produto a distribuição dessa variável é parecida com a anterior. Com média de 2280 dólares totais e 1394 de mediana é notável que a maioria dos clientes não possuem contratos de grande extensão. Para contratantes que saíram a média cai para 1531,8 dólares e mediana de 703,5 dólares. Para cobranças mensais o caminho é semelhante com média de \$ 64,76 e mediana de \$ 70.35 para todos os clientes e \$ 74.44 médios, com mediana em \$79,65, para clientes perdidos.

A figura 4.5 apresenta graficamente tais distribuições. Há evidência visual que cobranças mensais maiores tendem a aumentar a probabilidade de um cliente abandonar o produto.

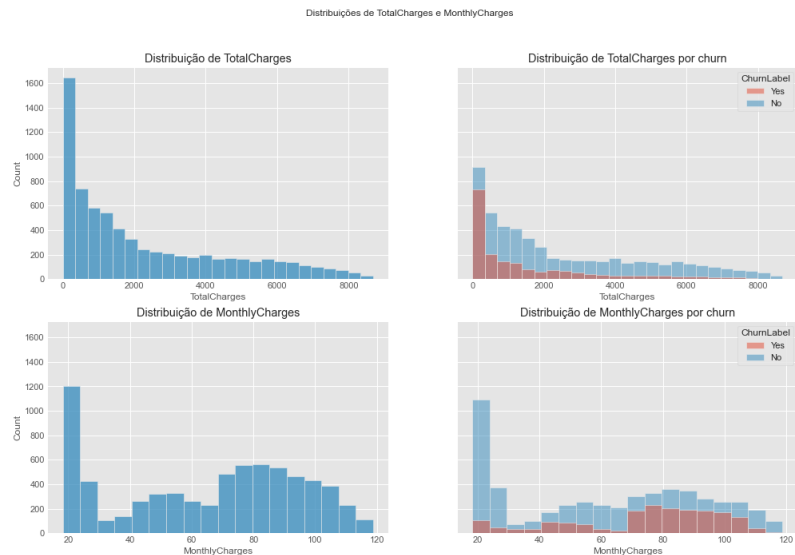


Figura 4.5: Distribuição de cobranças totais e por mês realizadas.

Terceiro, o uso médio da internet pode indicar a saída ou não do contratante. Assim, na base de dados, o valor médio em gigabytes é de 20,5, com uma mediana de 17 gigabytes utilizados em downloads. Para clientes que encerraram o contrato, essa média aumenta para 22,2 gigabytes, com uma mediana de 19 gigabytes. Evidentemente, clientes que abandonaram a plataforma tendem a utilizar mais o serviço de internet; isso pode indicar uma possível causa para o abandono, como velocidade não constante ou interrupções no serviço. A figura 4.6 apresenta essas distribuições.

Quarto, é válido apresentar a correlação entre cada variável numérica e categórica por meio de um mapa de calor, conforme mostrado na figura 4.7. Como demonstrado anteriormente, há uma relação positiva entre o tipo de contrato, especialmente contratos mais longos de um ou dois anos, e a quantidade de meses que o cliente permanece utilizando a plataforma. Por outro lado, há uma correlação negativa com o tipo de pagamento, especialmente cheques digitais. Essa análise fornece *insights* adicionais sobre as possíveis interdependências entre as variáveis e o *churn*.

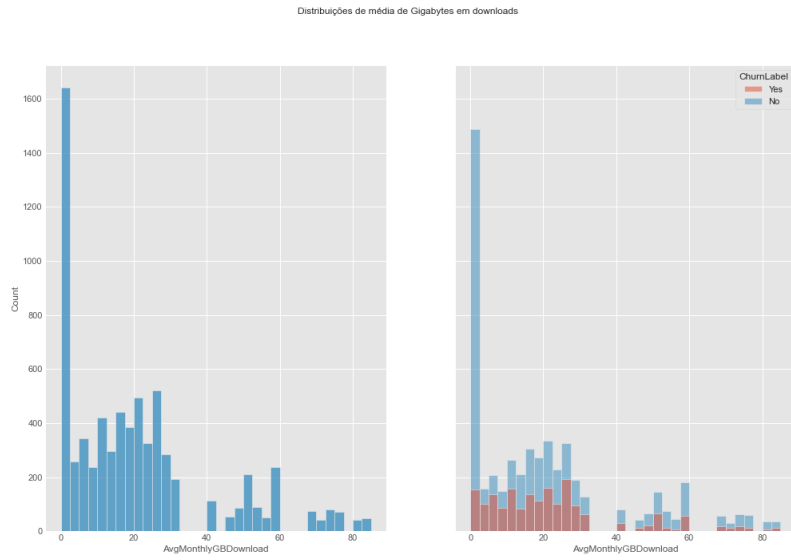


Figura 4.6: Distribuição de cobranças totais e por mês realizadas.



Figura 4.7: Correlação de Pearson entre as variáveis.

Por fim, é válido explorar a variável indicativa de *churn*, ChurnLabel. Há desbalanço entre a quantidade de clientes que deixaram a plataforma, apenas 26.5% do total, que pode dificultar a escolha do modelo de *machine learning* a ser escolhido. Contudo, é válido lembrar que há mais de 7000 observações garantindo a robustez do preditor. Além disso, o uso de metodologias de balanceamento dessa variável pode trazer possíveis falsas inferências, como exemplificado pelos autores de van den Goorbergh et al. (2022) para diversas técnicas em modelos de regressão logística. A figura 4.8 apresenta a distribuição do churn.

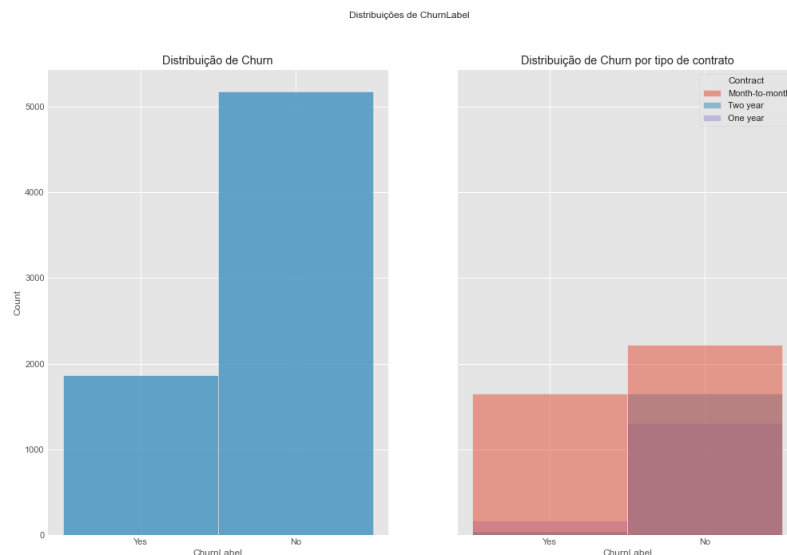


Figura 4.8: Distribuição da variável ChurnLabel e Distribuição estratificada por tipo de contrato

Manipulação de dados

As variáveis presentes na tabela 4.1 foram binarizadas de forma a proporcionar melhor inferência do modelo. Além disso, para o caso do descobrimento causal, análise de regressão, o modelo de contratos foi separado entre curta duração e longa duração. Variáveis numéricas foram padronizadas pelo processo presente no método de normalização "*StandardScaler*" da biblioteca "*Sklearn*" do Python. Essa normalização procura escalar

as variáveis de forma a manter a média zero e o desvio padrão igual a 1, $X \sim \mathcal{N}(0, 1)$.

4.1.2 Engenharia do modelo

Seguindo a literatura sobre o tema de previsão de *churn*, especialmente em Wu et al. (2021) e Rudd et al. (2021), os modelos de previsão de *churn* comumente utilizam o conceito de classificação para a identificação de clientes propensos a abandonar os serviços. Em essência, os algoritmos procuram dividir os clientes em duas classes, a variável alvo sendo a decisão de *churn* (abandono) ou não, e avaliam suas semelhanças por meio das variáveis de controle disponíveis.

A escolha de modelos de classificação é fundamentada na capacidade desses algoritmos de identificar padrões complexos nos dados e generalizá-los para fazer previsões em novos casos. A abordagem de classificação permite que o modelo classifique os clientes em categorias específicas, como "*churn*" ou "*não churn*", com base em características observadas.

Na engenharia de modelos voltados para a previsão de *churn* o processo é tradicionalmente dividido em cinco etapas cruciais. A primeira delas, a segmentação e preparação dos dados, abrange a identificação e segregação das variáveis pertinentes ao problema de *churn*, bem como a preparação dos dados para o treinamento do modelo. A segunda etapa, a escolha do modelo, compreende a seleção do algoritmo de classificação mais adequado ao problema, levando em consideração as características específicas dos dados e os requisitos do negócio. A terceira etapa, treinamento do modelo, que envolve a utilização de conjuntos de dados rotulados para ensinar o modelo a reconhecer padrões associados ao *churn*, ajustando seus parâmetros internos.

A otimização dos parâmetros -*hyperparameter tuning*-, quarta etapa, consiste no ajuste fino desses parâmetros para melhorar a performance do modelo. Esse refinamento muitas vezes é conduzido por meio de técnicas como busca em grade, *grid search*, ou otimização bayesiana. Finalmente, a quinta etapa é o teste, que se concentra na avaliação do desem-

penho do modelo em um conjunto de dados separado, conhecido como conjunto de teste. Desta maneira, permite verificar a capacidade de generalização do modelo e identificar possíveis problemas, como *overfitting*. O fluxograma de como todo esse caminho deve seguir é proposto pela figura 4.9.

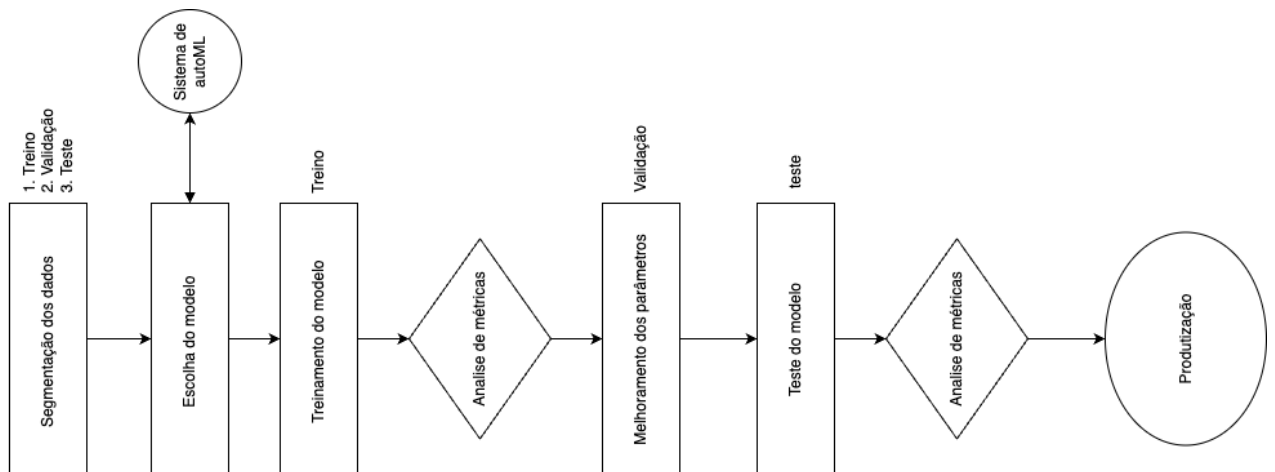


Figura 4.9: Fluxograma do modelo de aprendizado de máquina.

Segmentação e Preparação dos Dados

Os dados totais da base devem ser segmentados para garantir que o modelo aprenda padrões, seja otimizado e posto a prova com dados reais. Dessa forma, 60% da base de dados é dedicada a etapa de treinamento, 20% a de otimização de parâmetros e 20% a de testagem, sem reposição, dos dados já escolhidos. Adicionalmente, é válido pontuar que as transformações nos dados, citadas em 4.1.1 só devem ser realizadas após tal segmenta-

ção a fim de evitar o processo de vazamento de dados, "*data leakage*", entre as diferentes etapas podendo imputar viés ao modelo.

Escolha do modelo e Treinamento do modelo

Baseado em Wu et al. (2021), há diversos modelos para a escolha da melhor metodologia de previsão. Para tal, foi utilizado a biblioteca PyCaret de AutoML na escolha. O AutoML é um serviço no qual diversos modelos de aprendizado de máquina são testados sobre parte dos dados da pesquisa retornando os melhores algoritmos e modelos e suas métricas.

Os modelos que apresentaram melhor Score F1 e acurácia no experimento foram os baseados em LightGBM, técnica de menor custo computacional do uso do *gradiente boosting*. Assim, foi definido modelo de classificação baseado em LightGBM como padrão para a previsão do *churn*. O treinamento do modelo escolhido foi realizado com a base de treinamento e com os parâmetros padrões do LightGBM. Para a avaliação do modelo o mesmo foi rodado dez vezes para a análise de métricas como score F1, AUC e acurácia para que assim fosse escolhido o melhor modelo em média de todos os testados.

Melhoramento dos parâmetros do modelo, - *hyperparameter tuning*, e Teste

O melhoramento é realizado a partir da técnica de "*Grid Search*", o qual testa diversas combinações de padrões de parâmetros para a escolha do melhor modelo possível, desde regularizações em *L1* e *L2* até o conceito de amostragem. Para tal etapa, o melhoramento dos parâmetros foi feito com 20% da base total, diferentes da subseção anterior, afim de manter o menor viés. Por fim, foi produzido a modelagem de características ótimas para o problema analisado que está presente no *snippet* de código abaixo.

```
LGBMClassifier(boosting_type='gbdt', class_weight=None,
               colsample_bytree=1.0, importance_type='split',
               learning_rate=0.1, max_depth=-1,
```

```

min_child_samples=20, min_child_weight=0.001,
min_split_gain=0.0, n_estimators=100,
n_jobs=-1, num_leaves=31,
objective=None, reg_alpha=0.0, reg_lambda=0.00054,
subsample=1.0, subsample_for_bin=200000, subsample_freq=0)

```

Listing 4.1: Modelo utilizado como classificador

4.1.3 Análise causal

Considerando o conceito de causalidade apresentado por Angrist e Pischke (2015) e seguindo as orientações de modelagem delineadas por Chernozhukov et al. (2018) e Facure (2020), ela deve criar uma representação gráfica abrangente do processo causal analisado. No contexto específico do estudo, optou-se pelo estudo do efeito da duração de contratos no *churn*. É válido salientar que, devido à presença de mais de uma classe dentro da variável de tratamento (contrato), optou-se por selecionar uma para a análise causal. Dentro dos contratos, o contrato mês-a-mês foi escolhido como tratamento.

A variável a ser explicada, Y , foi definida como *ChurnLabel*, a variável de tratamento, T , como *Contract*, e o vetor de variáveis de controle, X cujo foram selecionadas a partir de escolhas por regressão *stepwise*, será analisado a partir do conceito de "*Double/Debiased Machine Learning*". Portanto, é crucial construir um grafo acíclico causal, conhecido como **DAG**, para visualizar de maneira clara as relações causais no contexto da intervenção e do fenômeno estudado. A relação entre as variáveis pode ser compreendida por meio da análise da intervenção, T , em relação à redução da incidência de *churn*, Y , enquanto as variáveis de controle fornecem uma explicação mais estratificada da causa geral do *churn*. A DAG que ilustra essas relações é apresentada na figura 4.10, oferecendo um modelo visual de causa e efeito relacionado à intervenção na política de preços ou continuidade do contrato.

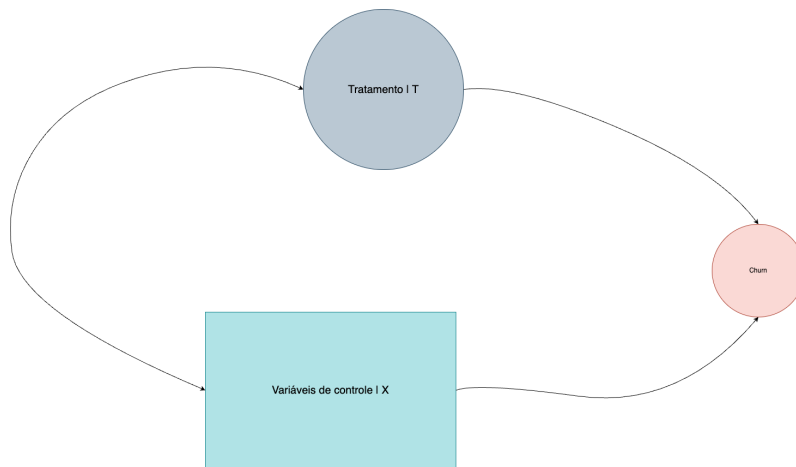


Figura 4.10: DAG do processo de análise causal.

Além disso, o modelo proposto em Chernozhukov et al. (2018) sugere a utilização de métodos de aprendizado de máquina para separar o efeito causal do tratamento das variáveis de controle. Dessa forma, na construção do modelo geral, foram selecionadas florestas aleatórias causais de regressão para ajustar a curva dos efeitos de controle na variável de tratamento e na variável explicada. Em seguida, para a regressão entre os resíduos de T e Y , utilizou-se uma regressão logística. Para realizar todos esses processos, foi utilizada a biblioteca econML, apresentada em Keith Battocchi (2019), do Python.

4.2 Qualidade do modelo, Entrega e Manutenção

No último ponto deste capítulo, é relevante destacar que um modelo de aprendizado de máquina é uma ferramenta passível de utilização em tomadas de decisão. A garantia de qualidade estabelece que boas decisões são baseadas em ferramentas eficazes, e, nesse sentido, optou-se por apresentar resultados as classes da variável ChurnLabel balanceadas e não balanceadas. Essa escolha visa manter um ponto de inferência válido, alinhando-se ao conceito apresentado no estudo de van den Goorbergh et al. (2022).

Além disso, as métricas de acurácia, score F1 e a curva ROC-AUC desempenham

papéis cruciais na avaliação e prosseguimento do modelo, sendo determinantes para assegurar sua qualidade. Por fim, optou-se pela utilização do método "*shapley additive explanations*", SHAP, conforme apresentado por Lundberg e Lee (2017), para investigar a importância de cada variável de controle e entender seu peso na decisão de cada observação em relação ao *churn* ou não. Essa abordagem proporciona uma compreensão mais aprofundada dos fatores que influenciam as previsões do modelo, contribuindo para a interpretabilidade e confiabilidade das decisões derivadas do sistema.

SHapley Additive exPlanations - SHAP

O método SHAP, ou "*shapley additive explanations*", desenvolvido por Lundberg e Lee (2017), é uma abordagem teórica que se baseia nos valores de Shapley da teoria dos jogos. Essa metodologia visa explicar as saídas de modelos de machine learning de maneira interpretável e equitativa. A ideia principal por trás do SHAP é atribuir contribuições individuais de cada variável para a predição final do modelo, considerando todas as combinações possíveis de características.

Ao calcular o valor de Shapley para uma característica específica, avalia-se a contribuição média dessa característica para todas as possíveis combinações de características, ponderadas pela sua importância relativa em cada combinação. Esse enfoque não apenas torna os modelos mais interpretáveis, mas também oferece uma abordagem justa e robusta para entender o impacto de cada variável nas previsões do modelo. Em cenários nos quais a transparência e a equidade são cruciais, o SHAP emerge como uma ferramenta valiosa para a interpretabilidade dos resultados do aprendizado de máquina. Um exemplo dessa abordagem pode ser vista na figura 4.11 onde o ganho de capital, a idade e a relação amorosa apresentam valores que podem interferir na predição da renda de adultos no Estados Unidos.

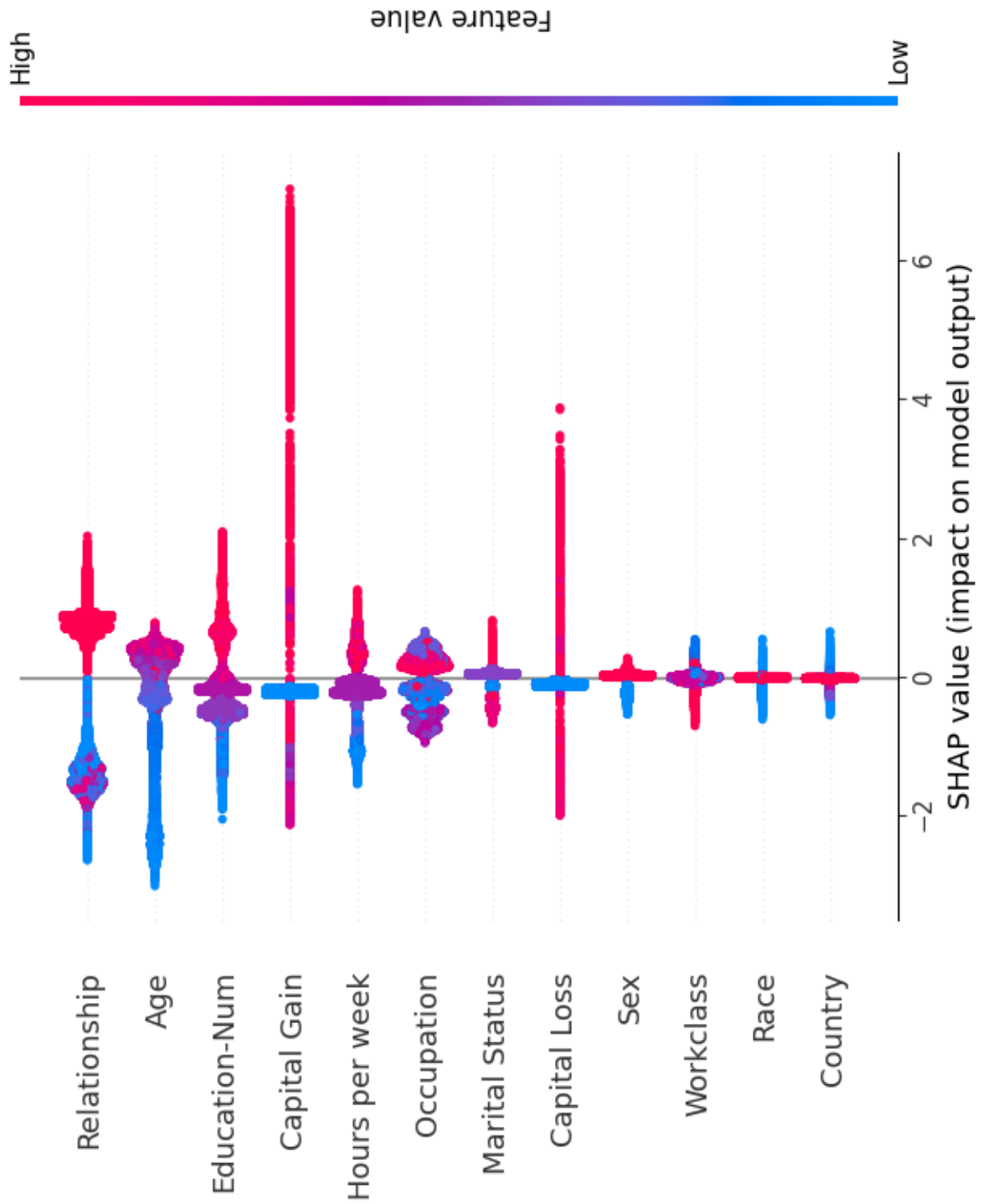


Figura 4.11: Exemplo do uso de SHAP em modelos de aprendizado de máquina.

Capítulo 5

Resultados

O modelo foi rodado seguindo os preceitos do CRISP-ML(Q) definido em Studer et al. (2020) para um classificador LGBM. A fim de manter o entendimento dos resultados este capítulo apresentará resultados para dois teste: um modelo não-balanceado e um modelo balanceado por "*oversampling*", assim como presente nos estudo dos autores de Wu et al. (2021) e Rudd et al. (2021). Dessa forma, o trabalho expõem como modelos calibrados de maneira errônea podem levar o tomador de decisão a resultados não cobertos pelo plano de ação como evidenciado por van den Goorbergh et al. (2022).

5.1 Métricas de avaliação - Classificador

Como descrito por Schneider (2016), para garantir que o modelo atende certos critérios de qualidade deve-se analisar suas métricas em relação de como ele classifica se algo é verdadeiro ou não e como este haje em relação a verdadeiros negativos.

Sendo assim, as métricas citadas na seção 2.4. Os resultados do score F1 e da Curva ROC-AUC foram utilizados como definidores do melhor modelo possível para o classificador.

Para o modelo sem o balanceamento da classe ChurnLabel, tem-se que a ferramenta classifica bem o que não é churn, tendo um valor de F1 de 0,84, e tem resultado satisfatório para a classificação do que é saída do cliente, F1 de 0,63 e ROC-AUC de 0,84. A

Tabela 5.1 expõem todas as métricas obtidas. Já o modelo balanceado apresenta melhores métricas em relação ao exposto anteriormente, como um escore F1 de 0,79 para Churn e ROC-AUC de 0,88. A Tabela 5.2 expõem todas as métricas obtidas.

Classe	Acurácia	Precisão	Sensibilidade	F1-Score	ROC-AUC
Não Churn (0)	-	0,82	0,83	0,84	-
Churn (1)	0,77	0,66	0,61	0,63	0,84

Tabela 5.1: Métricas classificador LGBM - Modelo não balanceado.

Classe	Acurácia	Precisão	Sensibilidade	F1-Score	ROC-AUC
Não Churn (0)	-	0,80	0,80	0,80	-
Churn (1)	0,81	0,80	0,79	0,79	0,88

Tabela 5.2: Métricas classificador LGBM - Modelo balanceado.

Ademais, a figura 5.1 demonstra a relação entre precisão e sensibilidade para cada ponto previsto pelo modelo. A Figura 5.3 apresenta a curva AUC-ROC para o modelo não balanceado. Para o modelo balanceado o mesmo contexto é encontrado nas figuras 5.2 e 5.4.

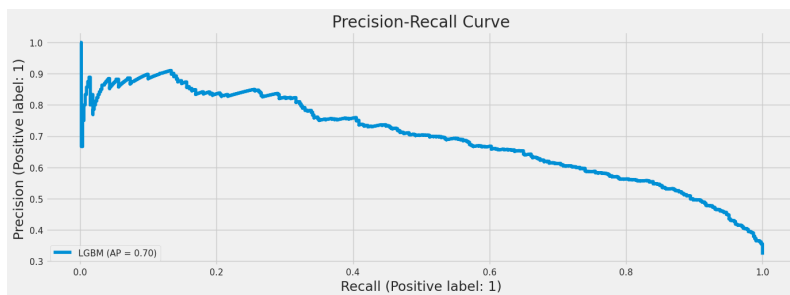


Figura 5.1: Gráfico de relação entre precisão e sensibilidade para o modelo sem balanceamento.

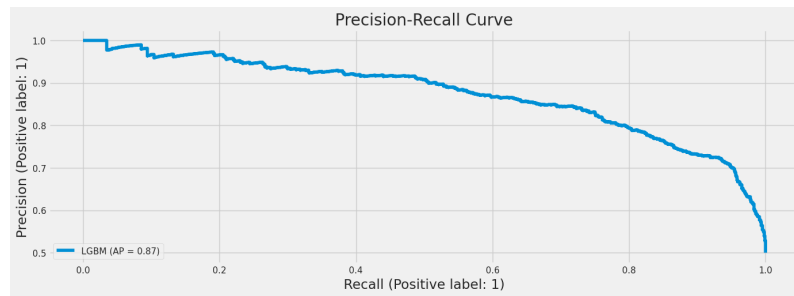


Figura 5.2: Gráfico de relação entre precisão e sensibilidade para o modelo com balanceamento.

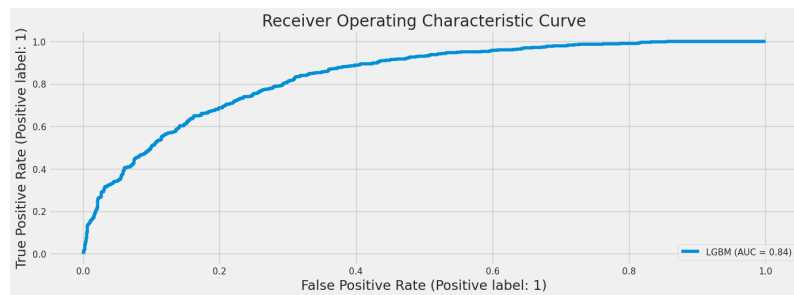


Figura 5.3: Gráfico da curva AUC-ROC para o modelo sem balanceamento.

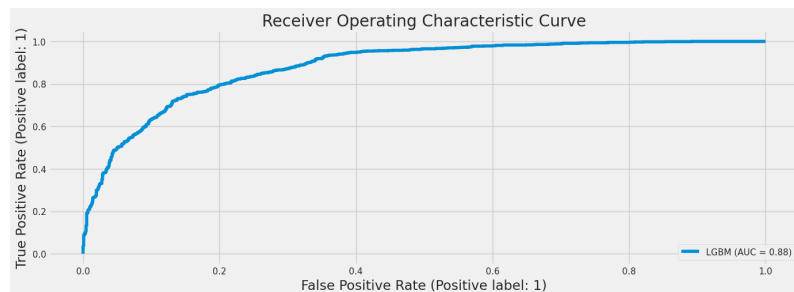


Figura 5.4: Gráfico da curva AUC-ROC para o modelo com balanceamento.

Os resultados seguem padrões encontrados em Wu et al. (2021) para o modelo de dados não balanceado; não obstante, o modelo de LGBM parece apresentar melhor sensibilidade, escore F1 e AUC-ROC que o apresentado no trabalho supra citado. Contudo, para o modelo com balanceamento ver-se que a uma deterioração nos valores de precisão e sensibilidade, por conseguinte, no escore F1 e uma manutenção de valores para o AUC-ROC.

Da avaliação do modelo a partir do que é apresentado por Hung et al. (2006), Huang et al. (2012), as variáveis de perfil de cliente e consumo tendem a guiar o classificador em conformidade com a literatura. O score F1 e AUC-ROC apresentam-se melhores, mas não devem ser comparados a risca devido a diferente design de projeto.

De forma continuada, os Gráficos de "uplift" e "wise uplift" são apresentados pelas Figuras 5.5 e 5.7. Tais métricas demonstram que o classificador, logo, o predictor são válidos para serem utilizados em tomadas de decisão referentes a o *churn*.

Há validação na literatura, por meio de Verhelst et al. (2020) e Verhelst et al. (2021a), de que os modelos balanceado ou não possuem bons resultados na classificação de clientes em churn. Os modelos definem que o ponto ótimo de separação, e classificação, é nas redondezas do quinto decil de ChurnLabel. O mesmo pode ser visto para o modelo com balanceamento, como evidenciado nas Figuras 5.6 e 5.8.

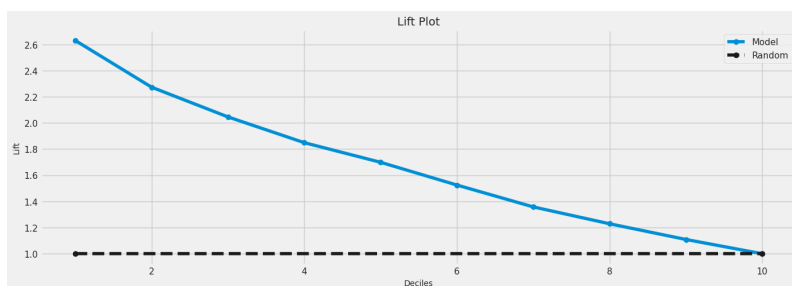


Figura 5.5: Gráfico de "uplift" para o modelo sem balanceamento.

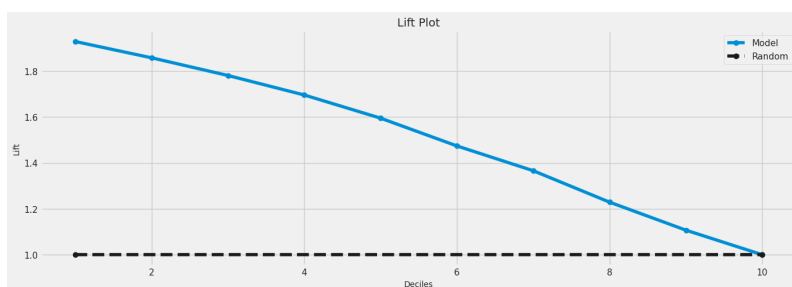


Figura 5.6: Gráfico de "uplift" para o modelo com balanceamento.

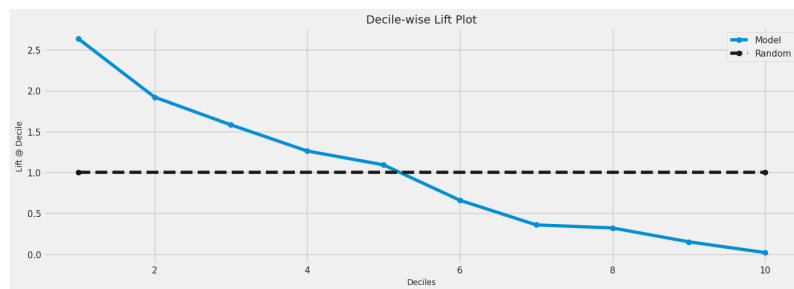


Figura 5.7: Gráfico de "wise uplift" para o modelo sem balanceamento.

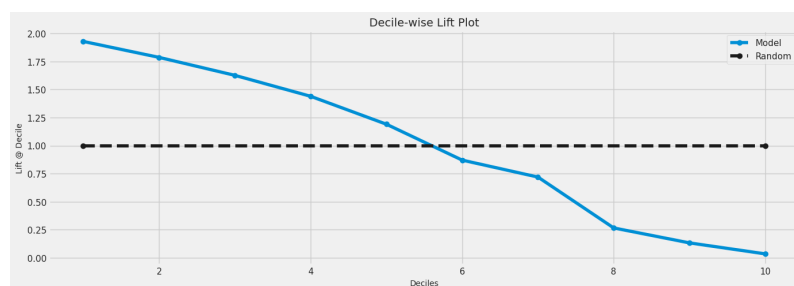


Figura 5.8: Gráfico de "wise uplift" para o modelo com balanceamento.

Para o gráfico cumulativo de ganho ver-se que o modelo de aprendizado de máquina é bem melhor que uma classificação aleatória e não apresenta "overfit" ao não sobrepor a linha "Wizard", cujo apresenta um modelo perfeito, como atestado em Facure (2020). De forma ingênua, pode-se inferir que o modelo com balanceamento tende a atingir o "tipping point" de "overfit" em valores menores que o modelo de base não balanceada. As Figuras 5.9 e 5.10 apresentam tal medida para o modelo sem balanceamento e com balanceamento, respectivamente.

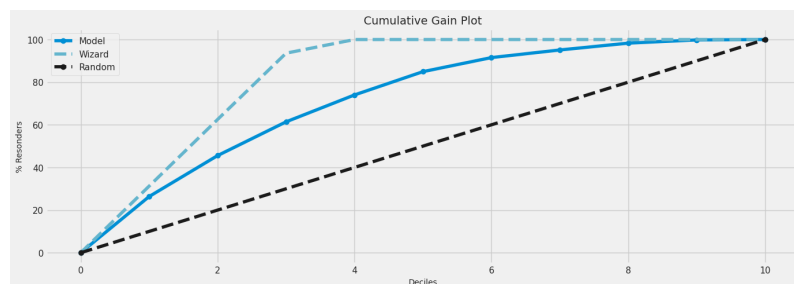


Figura 5.9: Gráfico de ganhos acumulados para o modelo sem balanceamento.

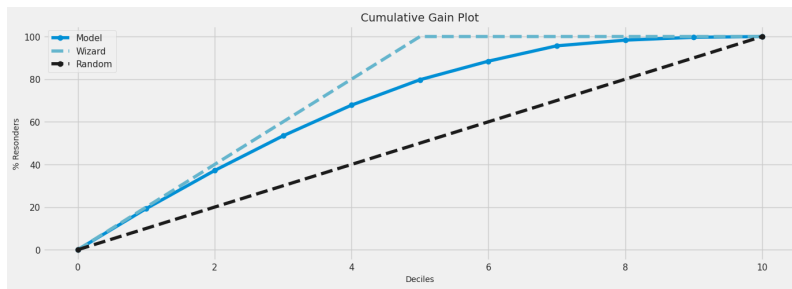


Figura 5.10: Gráfico de ganhos acumulados para o modelo com balanceamento.

Por fim, o modelo sem balanceamento apresenta por de corte ótimo no 5 decil com estatística KS de 51,4%, exposta na figura 5.11. O modelo com balanceamento apresenta o mesmo decil como ponto ótimo de corte mas com com estatística KS de 59,46%, exposta na figura 5.12.

Dessa forma, o valor da estatística ks entra em conformidade com as outras métricas ao evidenciar que o modelo atinge valores ótimos de separação nas redondezas do quinto decil.

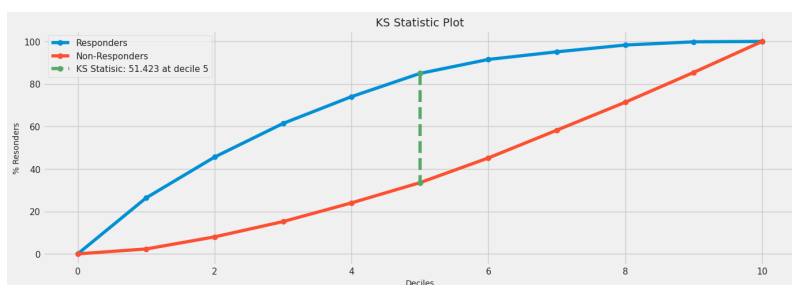


Figura 5.11: Estatística KS para o modelo sem balanceamento.

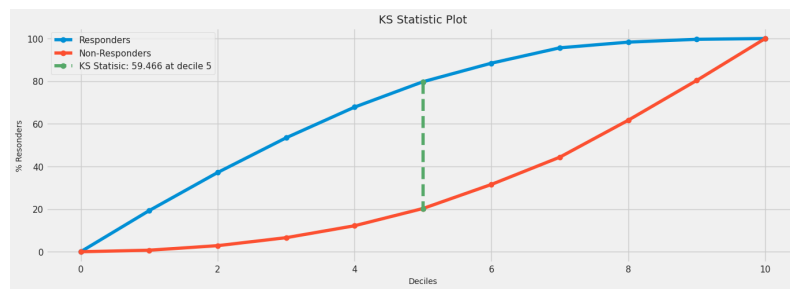


Figura 5.12: Estatística KS para o modelo com balanceamento.

5.2 SHAP Values e Importância das Variáveis

Nesta seção serão avaliadas a importância de cada variável para a identificação do *churn* pelo modelo LGBM. Para isso o conceito de SHAP será utilizado, descrito em Lundberg e Lee (2017), bem como a biblioteca presente no Python, para assim exibir graficamente as relações. Seguindo a apresentação do modelo não balanceado e balanceado.

Do resultado do "*feature importance*" do SHAP, tem-se que o ente de contratação mês-a-mês aliado a não presença de dependentes possuem grande influencia na desistência dos clientes perante o produto, como demonstrado na literatura por Wu et al. (2021). Adicionalmente, clientes com menor tempo de plataforma, menor valor de *TenureMonths*, e que opta pelo uso de fibra ótica também são mais propensos a sair. A Figura 5.13 apresenta a tal "*feature importance*" e como valores menores e maiores de cada variável dependem maior propensão ao churn.

O modelo com balanceamento apresenta importâncias de "*features*" diferentes. Para esta base de dados o custo alto de contas mensais, a presença de dependentes e serviços como suporte técnico e segurança online são de maior importância que o tipo de contrato ou a quantidade de meses que o cliente é usuário da plataforma.

O balanceamento por "*oversample*" pode ter alterado toda a distribuição de probabilidade de cada variável analisada. Assim, o conceito de tomada de decisão suportada por este modelo de aprendizado de máquina poderá ser comprometido, levando a dife-

rença entre os resultados e demonstrando empiricamente o que é descrito pelos autores de van den Goorbergh et al. (2022). A Figura 5.14 apresenta tais "features importances".

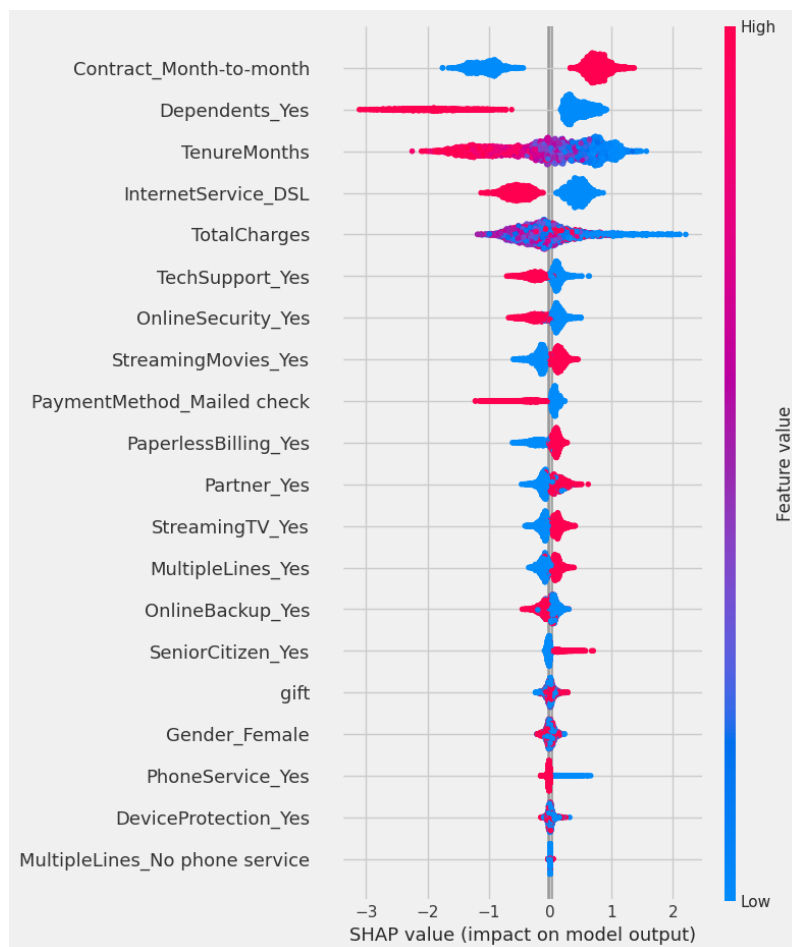


Figura 5.13: Gráfico de dispersão de importância de variáveis do SHAP para o modelo sem balanceamento.

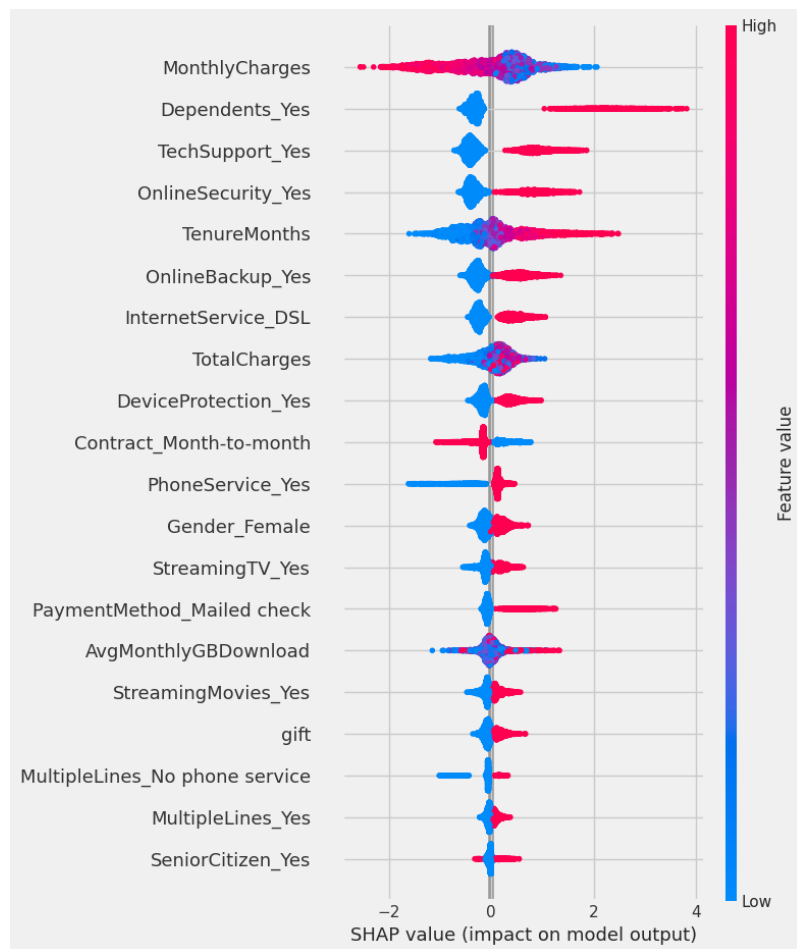


Figura 5.14: Gráfico de dispersão de importância de variáveis do SHAP para o modelo com balanceamento.

Da análise do modelo sem balanceamento, o gráfico apresentado na Figura 5.15 mostra que clientes com contrato mês-a-mês, que não possuem dependentes, apresentam maior propensão ao *churn*. Resultado similar é visto para clientes idosos na Figura 5.16, os quais apresentam maior chance de saída dado a presença de dependentes. Clientes que utilizam serviços de "streaming" de filmes tendem a permanecer mais na plataforma como visto na Figura 5.17. Ponto de inflexão para ser discutido é a presença de dependentes influi na saída do clientes tanto quanto a utilização desse serviço.

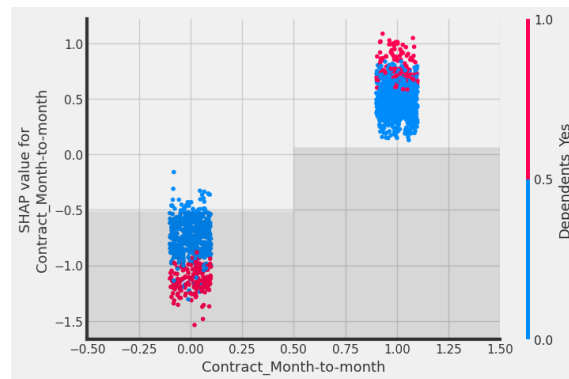
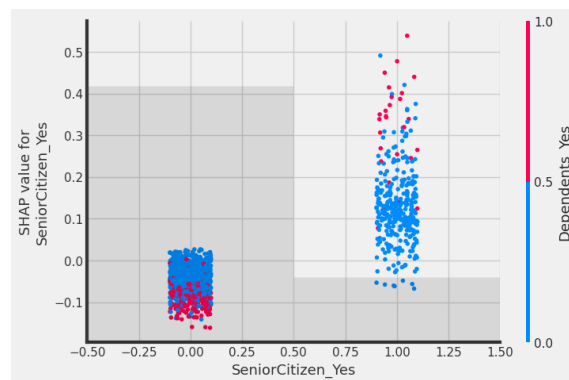
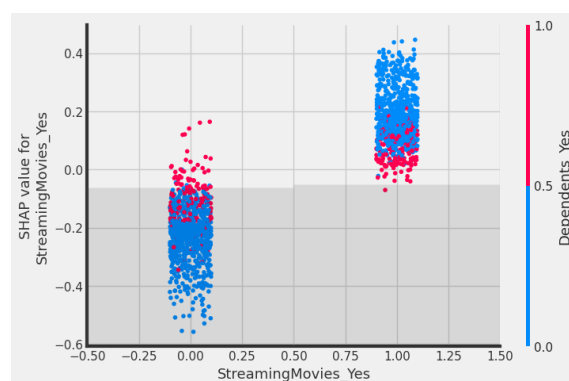
Figura 5.15: Distribuição *churn* por contrato.Figura 5.16: Distribuição *churn* por idade do cliente.

Figura 5.17: Distribuição Churn por uso de streaming de filmes.

Relações análogas podem ser vistas no modelo com balanceamento, mas é evidente que a distribuição dos valores SHAP é diferente, por vezes mais espaçada, e até magnitudes

trocadas como visto para a variável de *streaming* de filmes. As figuras 5.18, 5.19 e 5.20 apresentam essa relação.

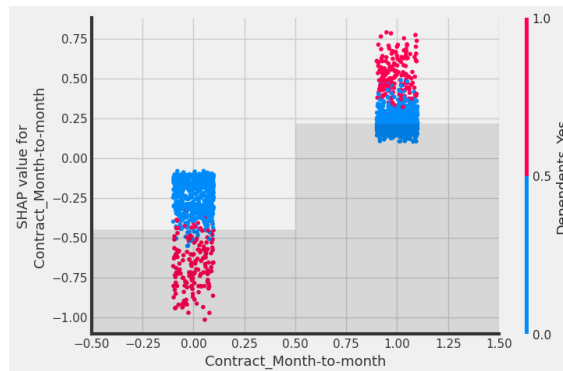


Figura 5.18: Distribuição *Churn* por contrato com base balanceada.

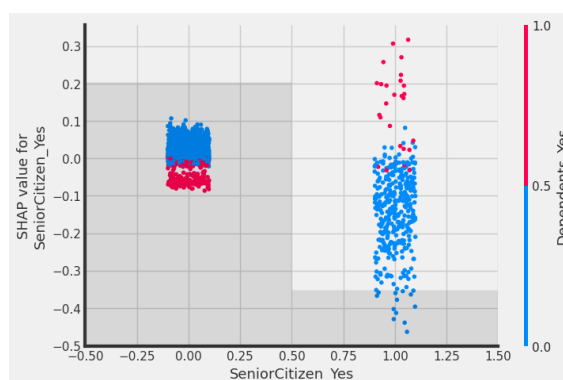


Figura 5.19: Distribuição *Churn* por idade do cliente com base balanceada.

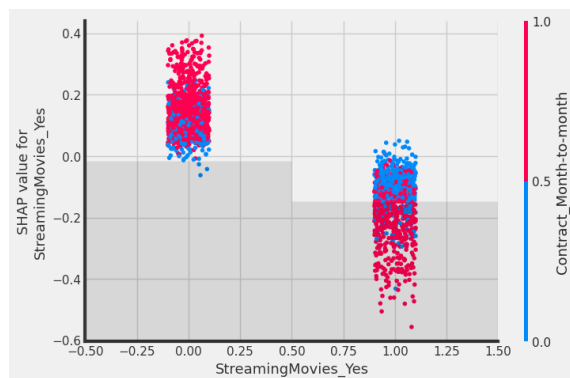


Figura 5.20: Distribuição *Churn* por uso de "streaming" de filmes com base balanceada.

A análise da quantidade de meses para o qual o contratante já possui na plataforma demonstra que sua permanência está quase que linearmente ligada a sua saída como é proposto por Wu et al. (2021). Usuários de diferentes tipos de contrato que possuem mais tempo de plataforma tendem a permanecer nela por mais tempo, para a base sem balanceamento. A Figura 5.21 evidencia essa relação.

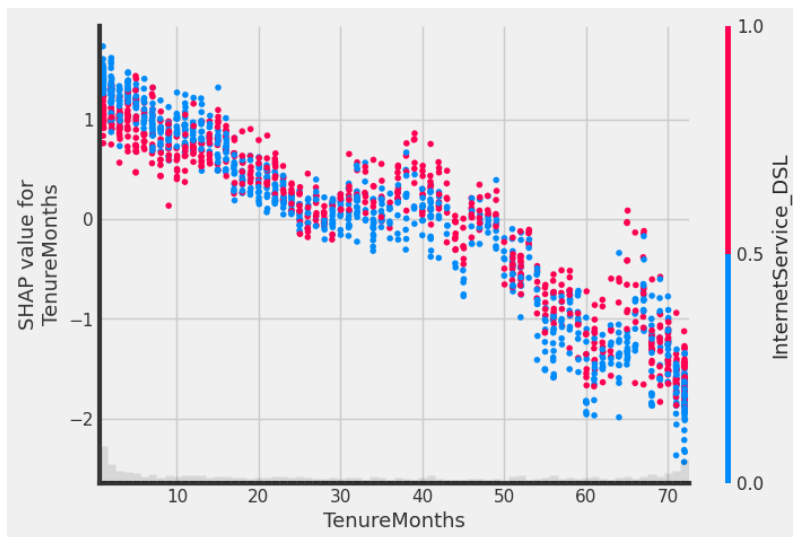


Figura 5.21: Distribuição *Churn* por tempo na plataforma.

Por fim, analisou-se a observação do banco de dados de teste para obter-se o porquê da saída ou permanência destes na utilização do produto contratado. Este usuário possui previsão de permanecer na plataforma com um valor de SHAP de -3,40 e uma esperança,

previsão, de -1,57. O fato de ele utilizar contrato de fidelização anual, modelo de internet DSL e permanência de muitos meses, 44, garante essa permanência. A Figura 5.22 apresenta um gráfico de decisão para tal usuário.

Para a base balanceada ver-se que as causas para qual o cliente permanece são diferentes, sendo a principal a contratação do serviço de suporte técnico e baixo custo mensal. A esperança -0,74 e valor SHAP de - 3,05 apresentando que o cliente está mais suscetível a sair da plataforma do que a realidade apresentada anteriormente. A Figura 5.23 demonstra a tomada de decisão do cliente.

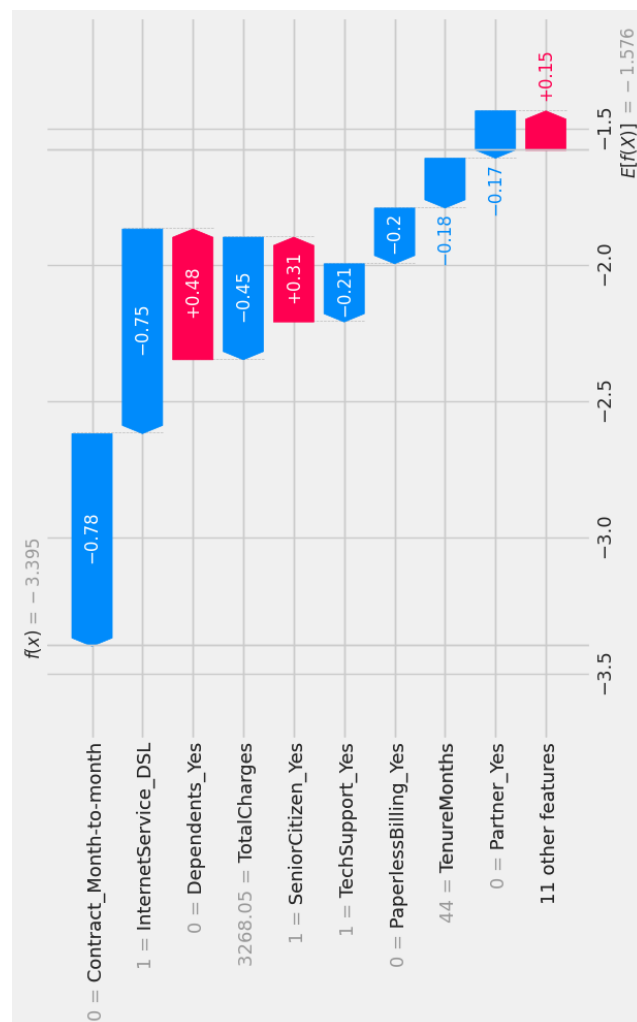


Figura 5.22: Análise de decisão do cliente por gráfico de cachoeira para modelo não balanceado.

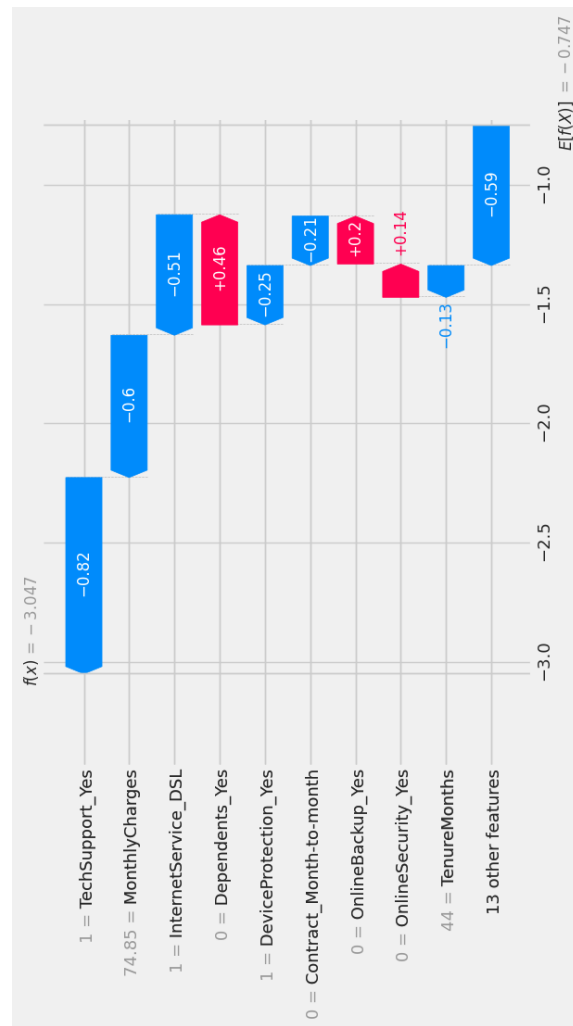


Figura 5.23: Análise de decisão do cliente por gráfico de cachoeira para modelo balanceado.

5.3 Análise Causal

A análise de diferenças significativas entre a causa do *churn* duração de contrato foi realizada a partir de uma regressão pelo método dos mínimos quadrados (OLS) e o método DML. A construção do DML seguiu as instruções citadas no capítulo anterior. Para a verificação das variáveis de tratamento foram construídas duas análises: regressão entre o *churn* e a variável de tratamento e outra contando com os controles. Além disso, a partir do resultado da regressão *stepwise* o vetor x de controle é formado pelas variáveis genero, método de pagamento, quantidade de meses como cliente e quantidade monetária total já

paga.

Da variável de duração de contratos, o tratamento contrato, tem-se que o tipo de contrato mês-a-mês apresenta uma relação de cerca de 37,2% de chance de abandonar a plataforma, com significância para $p < 0.001$, segundo a regressão (1), implicando que em análise simples tal conceito já evidencia sua importância. Além disso, o valor de R^2 de 0.152 denota que há relação causal evidente, mesmo que pouca, para esta relação e o valor da estatística F, de 989.220, depreende que há forte relação causal entre a saída do cliente e a duração do contrato.

Da análise da regressão com controles, é visto que há aumento 13,8% da chance de saída dos clientes devido a seu tipo de contrato como visto em (2). A diminuição de tal valor entre (2) e (1) é explicada por meio das variáveis de controle como pagamento por cheque eletrônico, que pode aumentar a saída em 12,2% e o uso de serviços de prioridade técnica que pode reduzir a saída em até 9,3%. Do R^2 ver-se uma melhor relação com a variável Y dado o valor de 0,222 e manutenção da relação causal das variáveis explicativas por meio da estatística F de 138.954. Tais valores podem ser vistos na Tabela 5.3.

Para a o resultado do efeito de tratamento condicional médio da regressão (3) o resultado é de 10%, com significância para $p < 0.001$, indicando que há relação causal direta entre os resíduos obtidos por DML. Contudo, é válido pontuar que o valor de R^2 é baixo, cerca de 0.01 e o valor da estatística F para (3) revela significância da regressão de forma global contudo com valores menores que (1) e (2). Da regressão com controles, (4) revela que não há causalidade de tais variáveis em relação a o *churn* de forma significativa. Porém, a valor significativo para os resíduos de contrato de cerca de 10%. Além disso, estatística F indica que mesmo que de forma diminuta há relação causal entre os resíduos do tipo de contrato e resíduos do *churn*.

Contudo, é visto que há a presença de heterocedasticidade em relação aos resíduos. A partir do teste de Goldfeld-Quandt tem-se valores de p-valor de 0,51 e estatística F de 0,99, pequena, corroborando para a rejeição da hipótese nula que não há variação crescente ou

	<i>Dependent variable: ChurnLabel</i>	
	(1)	(2)
contrato	0.372*** (0.012)	0.138*** (0.020)
Gender[T.Male]		0.009 (0.013)
PaymentMethod[T.Electronic check]		0.122*** (0.019)
Intercept	0.092*** (0.009)	0.432*** (0.039)
TechSupport[T.Yes]		-0.093*** (0.015)
TenureMonths		-0.010*** (0.001)
TotalCharges		-0.000 (0.000)
Observations	5517	5508
R^2	0.152	0.224
Adjusted R^2	0.148	0.222
Residual Std. Error	0.429 (df=5517)	0.410 (df=5508)
F Statistic	989.220*** (df=1; 5517)	138.954*** (df=8; 5508)

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Tabela 5.3: Regressões por OLS ChurnLabel e Contrato.

decrecente da variação dos resíduos. As regressões podem ser vistas na Tabela 5.4.

	<i>Dependent variable: churnResidual</i>	
	(3)	(4)
contrato res	0.100*** (0.020)	0.100*** (0.020)
Gender[T.Male]		0.002 (0.013)
PaymentMethod[T.Electronic check]		-0.005 (0.018)
Intercept	0.259*** (0.014)	0.269*** (0.023)
TechSupport[T.Yes]		-0.006 (0.014)
TenureMonths		0.000 (0.001)
TotalCharges		-0.000 (0.000)
Observations	3861	3852
R^2	0.010	0.007
Adjusted R^2	0.008	0.005
Residual Std. Error	0.392 (df=3859)	0.392 (df=3852)
F Statistic	25.458*** (df=1; 3859)	3.280*** (df=8; 3852)

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Tabela 5.4: Regressões por Random Forest e DML para churnResidual e ContratoResidual.

5.4 Análise de Resultados

Os resultados para os modelos de classificação em suas métricas de avaliação quando ao puro ato de separar diferentes conglomerados foram satisfatórios, principalmente para a base balanceada obtendo um escore F1 de 0,79 e uma curva AUC de 0,88 em comparação com os valores de 0,63 de escore F1 e 0,84 de curva AUC encontrados pelos autores de Wu et al. (2021). A Figura 5.24 apresenta tal comparação.

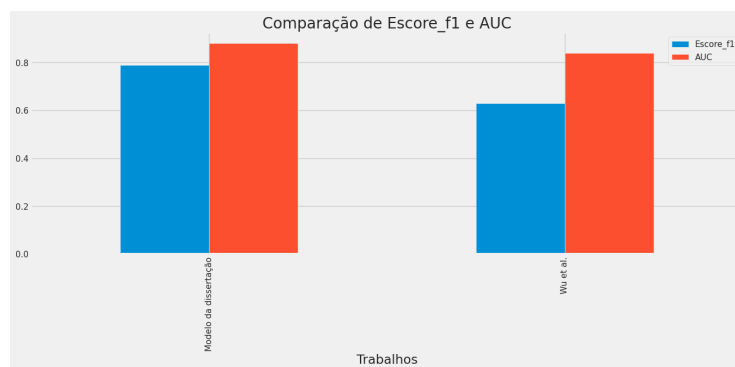


Figura 5.24: Comparação de resultados para base balanceada.

O modelo com base não balanceada também obteve boa performance com escore F1 de 0,63 e AUC-ROC de 0,84 cujo seguem em conformidade com o paragrafo anterior. Os resultados do estudo base, Wu et al. (2021), foram de 0,61 de escore F1 e 0,84 de AUC. A Figura 5.25 apresenta tal comparação.

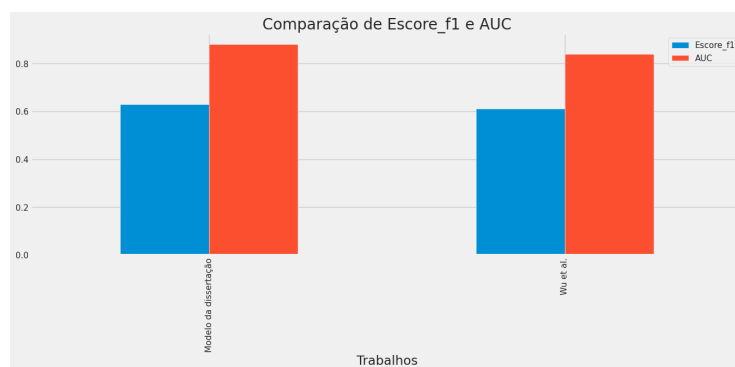


Figura 5.25: Comparação de resultados para base balanceada.

Os gráficos de uplift exibem que ambos os modelos performam bem para a classificação de *churn*, com wise uplift exibindo uma queda a partir do 5 decil, em relação a classificação aleatória seguindo o que é proposto pelos autores de Verhelst et al. (2020).

O resultado para o ganho acumulado demonstra que ambos os modelos não estão overfitados e podem ser utilizados em modelos de tomada de decisão. Do resultado da estatística KS, os resultados demonstram a melhor separação ocorre entre o 5 e o 6 decil. Tais resultados corroboram com os estudos já citados de Wu et al. (2021) e Verhelst et al. (2020) e acrescentam a definição de onde o classificador haze melhor.

O resultado dos valores SHAP demonstram concordância com o que é defendido pelos autores de van den Goorbergh et al. (2022). Os resultados para o modelo balanceado apresentam diferentes resultados de importância de variáveis que podem levar o tomador de decisão a buscar soluções não ótimas com a realidade. O fato de custos mensais e presença ou não de dependentes como entes de maior importância definem isso em comparação a maior importância do tipo de contrato no modelo não balanceado.

Do resultado da análise causal, é visto que hánexo causal entre o tipo de contrato e a *churn* corroborando com os resultados do classificador. Os resultados para a regressão linear também indicam que o tipo de pagamento e a contratação de serviços influencia no perca de clientes. Adicionalmente, para a análise de resíduos de contrato e churn por meio do DML é visto que há única relação forte é a dos resíduos do contrato para com o *churn*, porém por não respeitar o axioma da homoscedasticidade do teorema de Gauss-Markov não pode-se a inferir que tal estimador linear é o melhor estimador linear não viesado para a solução do problema. A figura 5.26 apresenta a relação crescente entre os resíduos e seus valores ajustados.

Os resultados expressaram concordância com o que é definido pelos autores de Devriendt et al. (2021) e Verhelst et al. (2021a), cujo defendem que o entendimento da razão da saída do cliente deve ser entendido como multifatorial e heterogêneo para a questão do uso de variáveis não residuais.

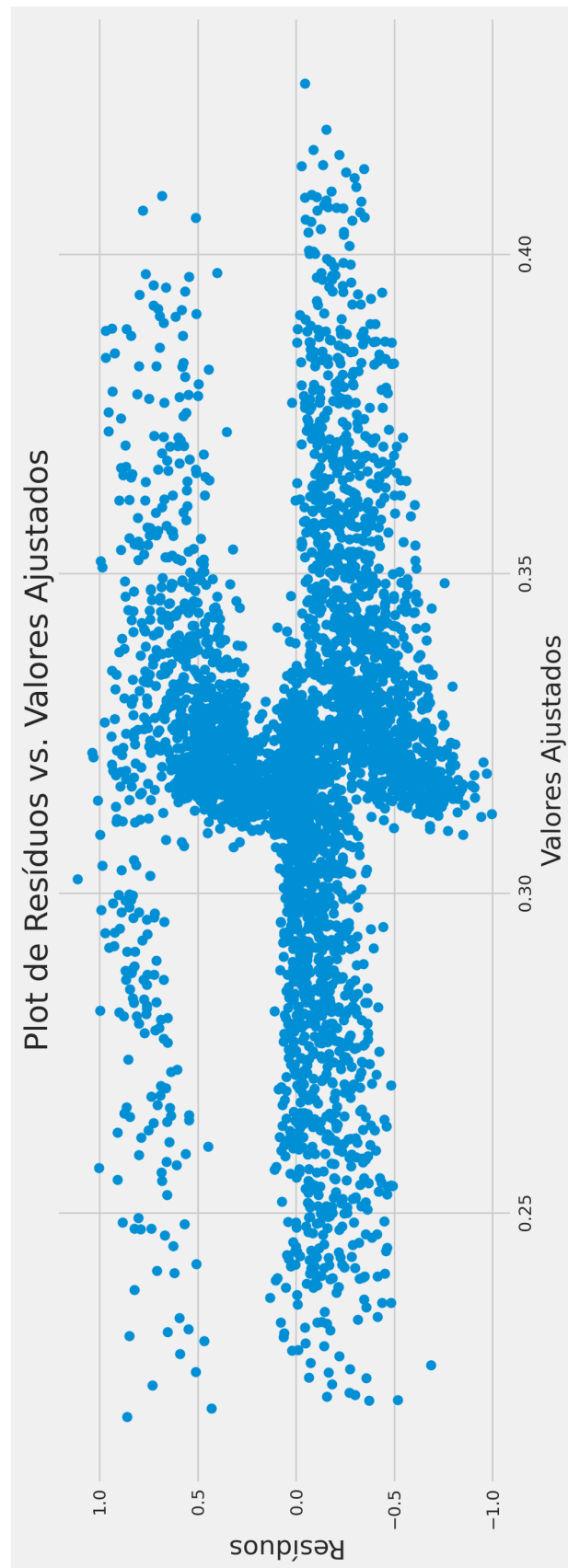


Figura 5.26: Relação entre os resíduos e seus valores ajustados.

Capítulo 6

Conclusão

Este trabalho teve como objetivo o desenvolvimento de dois modelos para o apoio de tomadas de decisão para conter a evasão de clientes. Neste contexto, a pesquisa buscou empreender sobre duas suposições para a base de dados utilizada. A primeira suposição define que a segmentação de clientes por características sociodemográficas, como idade, dependestes e gênero, pode definir maior índice de *churn*. A segunda suposição depreende que contratos de menor duração criam clientes menos leais a empresa.

Utilizando o trabalho dos autores de Wu et al. (2021) e Rudd et al. (2021) pode-se basear conceitos e *benchmarks* para constatação do exito do presente estudo. Além disso, objetivos específicos de qualidade e nível de acerto foram criados para que o desenvolvimento busca-se o melhor produto a ser entregue.

Do contexto estudado, o trabalho demonstra concordância com os trabalhos de base e literatura apresentada. Pode-se produzir modelo de previsão de saída com bons índices de qualidade e métricas de acerto, como exposto em Wu et al. (2021). Contudo, os resultados demonstram que variáveis relacionadas a contrato possuem maior importância que variáveis demográficas para o *churn*.

Adicionalmente, como já apresentado por Rudd et al. (2021), contratos de maior duração derivão em menores chances de perda de clientes. Porém, a utilização de uma técnica diferente, DML, do que é Rudd e co-autores não obteve resultados relevantes.

O trabalho também pode demonstrar que o balanceamento da variável alvo altera o

espaço da distribuição de probabilidades como evidenciando os resultados encontrados pelos autores de van den Goorbergh et al. (2022). Essa afirmação é validada pelos resultados de valores SHAP. Dessa forma, o estudo demonstra que o conceito de calibração de probabilidades deve ser respeitado para o atingimento de boas políticas de retenção pelo tomador de decisão.

Por fim, o presente estudo logrou êxito ao desenvolver uma ferramenta que apresenta um modelo de previsão de *churn*, baseado em Wu et al. (2021), que pode definir quais clientes estão próximos a abandonar o produto oferecido e pode usar o descobrimento causal, utilizando conceitos bases contidos em Rudd et al. (2021), para demonstrar as causas desta saída, como a duração do contrato influência na permanência do cliente. Do contexto de trabalhos futuros, a utilização de conceitos mais robustos como redes neurais para o descobrimento causal podem ajudar a melhorar o que foi encontrado por RUdd e seus co-autores e o presente estudo.

Referências Bibliográficas

Agrawal, Sanket, Aditya Das, Amit Gaikwad e Sudhir Dhage (2018), ‘Customer churn prediction modelling based on behavioural patterns analysis using deep learning’, *2018 International Conference on Smart Computing and Electronic Enterprise (ICS-CEE)* pp. 1–6.

URL: <https://api.semanticscholar.org/CorpusID:53950783>

Anderson, Chris (2008), *The Long Tail: Why the Future of Business Is Selling Less of More*, 2ª edição, Hyperion Books.

Angrist, Joshua D. e Jörn-Steffen Pischke (2008), *Instrumental Variables in Action: Sometimes You Get What You Need*, 1ª edição, Princeton University Press, p. 113–173.

Angrist, Joshua e Jörn-Steffen Pischke (2015), *Mastering 'Metrics: The path from cause to effect*, Princeton University Press.

Araújo, Jose Maria Amorim (2022), *Análise de Sobrevivência e Previsão de Churn de Clientes de Seguros de Vida do Banco do Brasil*, Tese de doutorado, Departamento de Ciência da Computação, Universidade de Brasília, Brasília, Brasil.

Ascarza, Eva (2018), ‘Retention Futility: Targeting High-Risk Customers Might be Ineffective’, *Journal of Marketing Research* **55**(1), 80–98.

URL: <https://doi.org/10.1509/jmr.16.0163>

Athey, Susan e Guido W Imbens (2019), ‘Machine Learning Methods That Economists Should Know About’, *Annual Review of Economics* **11**(1), 685–725.

URL: <https://doi.org/10.1146/annurev-economics-080217-053433>

Belcak, Peter e Roger Wattenhofer (2023), 'Fast feedforward networks'.

URL: <https://arxiv.org/pdf/2308.14711.pdf>

Botelho, Delane e Frederico Damian Tostes (2010), 'Modelagem de probabilidade de churn', *Revista de Administração de Empresas* **50**(4), 396–410.

URL: <https://doi.org/10.1590/S0034-75902010000400005>

Castanedo, Federico, Gabriel Valverde, Jaime Zaratiegui e Alfonso Vazquez (2014), Using deep learning to predict customer churn in a mobile telecommunication network.

URL: <https://api.semanticscholar.org/CorpusID:34331157>

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey e James Robins (2018), 'Double/debiased machine learning for treatment and structural parameters', *The Econometrics Journal* **21**(1), C1–C68.

Chiang, Ding-An, Yi-Fan Wang, Shao-Lun Lee e Cheng-Jung Lin (2003), 'Goal-oriented sequential pattern for network banking churn analysis', *Expert Systems with Applications* **25**(3), 293–302.

URL: <https://www.sciencedirect.com/science/article/pii/S0957417403000733>

Cunningham, Scott (2021), *Causal Inference: The Mixtape*, Yale University Press.

URL: <http://www.jstor.org/stable/j.ctv1c29t27>

Cutler, Adele, D Richard Cutler e John R Stevens (2012), *Random Forests*, Springer New York, New York, NY, pp. 157–175.

URL: https://doi.org/10.1007/978-1-4419-9326-7_5

de Franceschi, Pietro Reinheimer (2019), Modelagens preditivas de churn: o caso do banco do brasil, Tese de doutorado, Universidade do Vale do Rio dos Sinos - UNISINOS, Porto Alegre, Brasil.

Devriendt, Floris, Jeroen Berrevoets e Wouter Verbeke (2021), ‘Why you should stop predicting customer churn and start using uplift models’, *Information Sciences* **548**, 497–515.

URL: <https://www.sciencedirect.com/science/article/pii/S0020025519312022>

Dimitrova, Dimitrina, Vladimir Kaishev e Senren Tan (2020), ‘Computing the kolmogorov-smirnov distribution when the underlying cdf is purely discrete, mixed, or continuous’, *Journal of statistical software* **95**.

Duflo, Esther e Abhijit Banerjee (2017), *Handbook of Field Experiments*, North Holland.

Facure, Matheus (2020), Causal inference for the brave and true, Causal inference and causal machine learning manual, Nubank, Brazil.

URL: <https://matheusfacure.github.io/python-causality-handbook/landing-page.html>

Geron, Aurelien (2019), *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2ndª edição, O’Reilly Media, Inc.

Ghorbani, Aminah e Fattaneh Taghiyareh (2009), Cmf: A framework to improve the management of customer churn, em ‘2009 IEEE Asia-Pacific Services Computing Conference (APSCC)’, pp. 457–462.

Huang, Bingquan, Mohand Tahar Kechadi e Brian Buckley (2012), ‘Customer churn prediction in telecommunications’, *Expert Systems with Applications* **39**(1), 1414–1425.

URL: <https://www.sciencedirect.com/science/article/pii/S0957417411011353>

Hung, Shin-Yuan, David C Yen e Hsiu-Yu Wang (2006), ‘Applying data mining to telecom churn management’, *Expert Systems with Applications* **31**(3), 515–524.

URL: <https://www.sciencedirect.com/science/article/pii/S0957417405002654>

Izza, Yacine, Alexey Ignatiev e João Marques-Silva (2020), ‘On explaining decision trees’, *CoRR abs/2010.11034*.

URL: <https://arxiv.org/abs/2010.11034>

J. Burez, D. Van den Poel (2009), ‘Handling class imbalance in customer churn prediction’, *Expert Systems with Applications* **36**(3, Part 1), 4626–4636.

URL: <https://www.sciencedirect.com/science/article/pii/S0957417408002121>

Jaffery, T. S. e Shirley Liu (2009), Measuring campaign performance by using cumulative gain and lift chart.

URL: <https://api.semanticscholar.org/CorpusID:8287576>

Keith Battocchi, Eleanor Dillon, Maggie Hei Greg Lewis Paul Oka Miruna Oprescu Vasilis Syrgkanis (2019), ‘EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation’, <https://github.com/py-why/EconML>. Version 0.x.

Kurtz, David L. e Kenneth E. Clow (1998), *Services marketing*, 1ª edição, John Wiley Sons, New York, EUA.

Künzel, Sören R., Jasjeet S. Sekhon, Peter J. Bickel e Bin Yu (2019), ‘Metalearners for estimating heterogeneous treatment effects using machine learning’, *Proceedings of the National Academy of Sciences* **116**(10), 4156–4165.

URL: <http://dx.doi.org/10.1073/pnas.1804597116>

Lundberg, Scott M e Su-In Lee (2017), A unified approach to interpreting model predictions, em ‘Advances in Neural Information Processing Systems 30’, Curran Associates, Inc., pp. 4765–4774.

URL: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>

Mehlig, Bernhard (2019), ‘Artificial neural networks’, *CoRR abs/1901.05639*.

URL: <http://arxiv.org/abs/1901.05639>

Mishra, Abinash e U. Srinivasulu Reddy (2017), ‘A novel approach for churn prediction using deep learning’, *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)* pp. 1–4.

URL: <https://api.semanticscholar.org/CorpusID:53280048>

Molak, Aleksander (2023), *Causal Inference and Discovery in Python: Unlock the secrets of modern causal machine learning with DoWhy, EconML, PyTorch and more*, 1.^a edição, Packt Publishing, Birmingham. <https://amzn.to/3RebWzn>.

Natekin, Alexey e Alois Knoll (2013), ‘Gradient boosting machines, a tutorial’, *Frontiers in neurorobotics* **7**, 21.

Neslin, Scott A, Sunil Gupta, Wagner Kamakura, Junxiang Lu e Charlotte H Mason (2006), ‘Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models’, *Journal of Marketing Research* **43**(2), 204–211.

URL: <https://doi.org/10.1509/jmkr.43.2.204>

Pinheiro, Paulo e Luís Cavique (2022), Telco Customer Churn Analysis: Measuring the Effect of Different Contracts, em A.Rocha, H.Adeli, G.Dzemyda e F.Moreira, eds., ‘Information Systems and Technologies’, Springer International Publishing, Cham, pp. 112–121.

Quinlan, J R (1986), ‘Induction of decision trees’, *Machine Learning* **1**(1), 81–106.

URL: <https://doi.org/10.1007/BF00116251>

Rudd, David Hason, Huan Huo e Guandong Xu (2021), Causal analysis of customer churn using deep learning, em ‘2021 International Conference on Digital Society and Intelligent Systems (DSInS)’, pp. 319–324.

Rumelhart, David E, Geoffrey E Hinton e Ronald J Williams (1986), ‘Learning representations by back-propagating errors’, *Nature* **323**(6088), 533–536.

URL: <https://doi.org/10.1038/323533a0>

Schneider, Pedro Henrique (2016), *Análise preditiva de Churn com ênfase em técnicas de Machine Learning: Uma Revisão*, Tese de doutorado, EMap, Fundação Getulio Vargas, Rio de Janeiro, Brasil.

Semenova, Vira, Matt Goldman, Victor Chernozhukov e Matt Taddy (2017), *Estimation and inference on heterogeneous treatment effects in high-dimensional dynamic panels*, Working Paper 8456, MIT.

URL: <https://arxiv.org/abs/1712.09988>

Studer, Stefan, Thanh Binh Bui, Christian Drescher, Alexander Hanuschkin, Ludwig Winkler, Steven Peters e Klaus-Robert Müller (2020), ‘Towards CRISP-ML(Q): A machine learning process model with quality assurance methodology’, *CoRR abs/2003.05155*.

URL: <https://arxiv.org/abs/2003.05155>

Swift, Ronald S. (2001), *Accelerating Customer Relationships: Using Crm and Relationship Technologies*, 1ª edição, Prentice Hall.

van den Goorbergh, Ruben, Maarten van Smeden, Dirk Timmerman e Ben Van Calster (2022), ‘The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression’.

Veloso, Fernando Jorge Machado (2012), *Um Modelo para Previsão de Churn na Área do Retalho*, Tese de doutorado, Escola de Engenharia, Universidade de Minho, Minho, Portugal.

Verhelst, Théo, Jeevan Shrestha, Denis Mercier, Jean-Christophe Dewitte e Gianluca Bontempì (2021a), *Predicting Reach to Find Persuadable Customers: Improving Uplift Models for Churn Prevention*, em C.Souares e L.Torgo, eds., ‘Discovery Science’, Springer International Publishing, Cham, pp. 44–54.

- Verhelst, Théo, Jeevan Shrestha, Denis Mercier, Jean-Christophe Dewitte e Gianluca Bontempi (2021*b*), Predicting Reach to Find Persuadable Customers: Improving Uplift Models for Churn Prevention, *em* C.Soares e L.Torgo, eds., ‘Discovery Science’, Springer International Publishing, Cham, pp. 44–54.
- Verhelst, Théo, Olivier Caelen, Jean-Christophe Dewitte, Bertrand Lebichot e Gianluca Bontempi (2020), Understanding Telecom Customer Churn with Machine Learning: From Prediction to Causal Inference, *em* ‘Artificial Intelligence and Machine Learning’, Springer International Publishing, Cham, pp. 182–200.
- Wager, Stefan e Susan Athey (2018), ‘Estimation and Inference of Heterogeneous Treatment Effects using Random Forests’, *Journal of the American Statistical Association* **113**(523), 1228–1242.
URL: <https://doi.org/10.1080/01621459.2017.1319839>
- Wooldridge, Jeffrey Marc (2016), *Introdução à Econometria: Uma Abordagem Moderna.*, Vol. 4, Cengage Learning., São Paulo, capítulo 17 - Modelos com variáveis dependentes limitadas e correções da seleção amostral, pp. 647–689.
- Wu, Shuli, Wei-Chuen Yau, Thian Song Ong e Siew-Chin Chong (2021), ‘Integrated churn prediction and customer segmentation framework for telco business’, *IEEE Access* **9**, 62118–62136.
URL: <https://api.semanticscholar.org/CorpusID:233434157>
- Xie, Yaya, Xiu Li, E W T Ngai e Weiyun Ying (2009), ‘Customer churn prediction using improved balanced random forests’, *Expert Systems with Applications* **36**(3, Part 1), 5445–5449.
URL: <https://www.sciencedirect.com/science/article/pii/S0957417408004326>