

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROJETO DE GRADUAÇÃO**

BRENO DINIZ CORREA LIMA

**PREVISÃO DE CANCELAMENTO DE ASSINATURAS DE
VINHO COM TÉCNICAS DE APRENDIZADO DE
MÁQUINA**

VITÓRIA
2021

BRENO DINIZ CORREA LIMA

**PREVISÃO DE CANCELAMENTO DE ASSINATURAS DE VINHO
COM TÉCNICAS DE APRENDIZADO DE MÁQUINA**

Parte manuscrita do Projeto de Graduação do aluno **Breno Diniz Correa Lima**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Orientador: Prof. Dr. Helder Roberto de Oliveira Rocha

Coorientador: M.Sc Higor Araújo Fim Camporez

VITÓRIA
2021

BRENO DINIZ CORREA LIMA

**PREVISÃO DE CANCELAMENTO DE ASSINATURAS DE VINHO
COM TÉCNICAS DE APRENDIZADO DE MÁQUINA**

Parte manuscrita do Projeto de Graduação do aluno **Breno Diniz Correa Lima**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Aprovada em 13 de Outubro de 2021.

COMISSÃO EXAMINADORA:

Helder Roberto de O. Rocha

Prof. Dr. Helder Roberto de Oliveira Rocha
Universidade Federal do Espírito Santo
Orientador

Higor A. F. Camporez

M.Sc. Higor Araújo Fim Camporez
Universidade Federal do Espírito Santo
Coorientador

Jair Adriano Lima Silva

Prof. Dr. Jair Adriano Lima Silva
Universidade Federal do Espírito Santo
Examinador

Wesley da Silva Costa

M.Sc. Wesley da Silva Costa
Universidade Federal do Espírito Santo
Examinador

Aos meus pais, Luciano e Sandra.

AGRADECIMENTOS

Agradeço primeiramente à Deus pela oportunidade de realizar este trabalho.

Aos meus pais, Luciano Diniz e Sandra Luiza, por toda a educação que me proporcionaram e apoio para que conseguisse concluir mais essa etapa na minha vida.

À toda minha família, que sempre esteve ao meu lado e é a base para que eu alcance os meus objetivos.

À todos os amigos que fiz durante essa caminhada, trouxeram muito aprendizado e alegria à estes anos e fizeram da Universidade uma segunda casa para mim.

Ao meu orientador Prof. Dr. Helder Roberto de Oliveira Rocha e coorientador M.Sc. Higor Araújo Fim Camporez por todo apoio, direcionamento e conhecimento passado.

À todos os professores e servidores da UFES, UFPE e UAlg por onde passei por estes anos de graduação. Obrigado por fazerem a diferença no nosso país.

Agradeço também à empresa Wine e à todos os colaboradores por todo apoio, incentivo, acolhimento e dedicação neste trabalho.

RESUMO

Planos de assinaturas fornecem produtos ou serviços periodicamente por meio de cobranças automáticas, por exemplo, empresas de telecomunicações fornecem mensalmente serviços como ligações e internet e empresas de assinaturas de vinho entregam vinho mensalmente por um certo valor pago no início do mês. Por vários motivos, eles representam uma tendência de modelo de negócio para as empresas. Do lado do assinante, a praticidade é o grande diferencial, não é necessário realizar a compra nem gastar tempo para escolher o produto, a compra é efetuada automaticamente e os produtos são escolhidos uma só vez no momento da assinatura geralmente classificados em categorias por perfil de consumidor. Do lado das empresas, a previsibilidade e a garantia de receita são as maiores vantagens, com planos de assinaturas é mais fácil calcular a receita do próximo mês por exemplo, além disso, essa receita é mais palpável do que receitas de vendas avulsas. Sendo assim, a única ameaça às vantagens apresentadas para as empresas é o cancelamento de assinaturas. Portanto, este projeto visa o desenvolvimento de modelos de aprendizado de máquina para prever quantitativamente e qualitativamente os cancelamentos de assinaturas mensais de vinho da empresa Wine. A previsão quantitativa irá indicar quantos assinantes irão cancelar as assinaturas no mês seguinte por meio de modelos supervisionados de regressão, *Random Forest* e rede neural *Feed Forward* em conjunto com LSTM a fim de descobrir qual possui melhor performance para o problema. Já a previsão qualitativa, irá indicar quais assinaturas serão canceladas no mês seguinte por meio de dois modelos supervisionados de classificação, *Random Forest* e rede neural *Feed Forward*.

Palavras-chave: Assinaturas mensais. Vinho. Aprendizado de máquina.

ABSTRACT

Subscription plans provide products or services periodically through automatic billing, for example, telecommunications companies provide monthly services such as calls and internet and wine subscription companies deliver wine monthly for a certain amount paid at the beginning of the month. For several reasons, they represent a business model trend for companies. On the subscriber's side, practicality is the big difference, there is no need to make a purchase or spend time to choose the product, the purchase is made automatically and the products are chosen once at the time of subscription, usually classified into categories by profile. On the business side, predictability and guaranteed revenue are the biggest advantages, with subscription plans it's easier to calculate next month's revenue for example, in addition, this revenue is more palpable than revenue from single sales. Therefore, the only threat to the advantages presented to companies is the cancellation of subscriptions. Therefore, this project aims to develop machine learning models to quantitatively and qualitatively predict the monthly wine subscription cancellations of the Wine company. The quantitative forecast will indicate how many subscribers will unsubscribe in the following month through two supervised regression models, Random Forest and Feed-Forward neural network with LSTM, in order to find out which one performs better for the problem. The qualitative forecast will indicate which subscriptions will be canceled in the following month through two supervised classification models, Random Forest and Feed-Forward neural network.

Keywords: Subscription plan. Wine. Machine Learning.

LISTA DE FIGURAS

Figura 1 – Estrutura da árvore de decisão	20
Figura 2 - Esquema de ligação entre os neurônios no cérebro humano.	23
Figura 3 - Esquema de funcionamento de um neurônio	24
Figura 4 - Rede neural <i>Feed Forward</i>	26
Figura 5 – LSTM.	28
Figura 6 – Mecanismo de esquecimento da LSTM.	28
Figura 7 – Mecanismo de armazenamento da LSTM.....	29
Figura 8 – Mecanismo de atualização do neurônio da LSTM.....	29
Figura 9 – Mecanismo de saída do neurônio da LSTM.....	30
Figura 10 - Extração dos dados	32
Figura 11 - Estrutura da rede neural <i>Feed Forward</i> com LSTM	48
Figura 12 - Estrutura rede neural <i>Feed Forward</i>	54

LISTA DE GRÁFICOS

Gráfico 1 – Funções de ativação: (a) função tangente hiperbólica, (b) função sigmóid e (c) função rampa	25
Gráfico 2 - Distribuição do conjunto de dados para modelos de predição quantitativa.	46
Gráfico 3 - Valores reais e estimados para conjunto de dados de teste - <i>Random Forest</i>	47
Gráfico 4 - MAPE - <i>Feed Forward</i> com LSTM.....	49
Gráfico 5 - Valores reais e estimados para o conjunto de teste - <i>Feed Forward</i> com LSTM	50
Gráfico 6 - Distribuição dos dados predição qualitativa.	51
Gráfico 7 - Conjunto de dados de treinamento.	52
Gráfico 8 - Conjunto de dados de teste.....	52

LISTA DE QUADROS

Quadro 1 - Atributos do conjunto de dados da predição quantitativa	36
Quadro 2 - Atributos do conjunto de dados da predição qualitativa	37
Quadro 3 – Parâmetros para a técnica <i>Random Forest</i>	46
Quadro 4 – Média percentual do erro absoluto – <i>Random Forest</i>	47
Quadro 5 – Média percentual do erro absoluto – <i>Feed Forward</i> com LSTM.....	49
Quadro 6 – Comparação do desempenho das técnicas	50
Quadro 7 – Parâmetros pra a técnica <i>Random Forest</i>	53
Quadro 8 - Acurácia para conjunto de treinamento e teste - <i>Random Forest</i>	53
Quadro 9 - Matriz confusão - <i>Random Forest</i>	54
Quadro 10 - Acurácia para conjunto de treinamento e teste - <i>Feed Forward</i>	55
Quadro 11 - Matriz de confusão - <i>Feed Forward</i>	55
Quadro 12 - Comparação de desempenho das técnicas.....	56

LISTA DE ABREVIATURAS E SIGLAS

Fp	Falsos positivos
Fn	Falsos Negativos
LSTM	<i>Long Short-Term Memory</i>
MAPE	Média Percentual do Erro Absoluto
RNN	<i>Recurrent Neural Network</i>
UFES	Universidade Federal do Espírito Santo
Vp	Verdadeiros positivos
Vn	Verdadeiros negativos

LISTA DE SÍMBOLOS

C	Estado do neurônio
d	Valor original no conjunto de dados
Gini	Coefficiente de Gini
h	Saída do neurônio
k	Quantidade de classes em um conjunto de dados
n	Quantidade de dados no conjunto de dados de treinamento
m	Número de neurônios
O	Função de saída do neurônio
p	Probabilidade relativa
SN	Saída do neurônio
tanh	Função tangente hiperbólica
V	Variância
U	Valor normalizado no conjunto de dados
y	Saída do conjunto de dados
y_m	Valor médio da saída do conjunto de dados
w	Pesos aplicados às entradas do neurônio
z	Valor total recebido pelo neurônio
Φ	Função Sigmóid

SUMÁRIO

1	INTRODUÇÃO.....	15
	1.1 Apresentação e Objeto de Pesquisa	15
	1.2 Justificativa	16
	1.3 Objetivos.....	17
	1.3.1 Objetivo geral	17
	1.3.2 Objetivos Específicos	17
	1.4 Estrutura do Texto	17
2	EMBASAMENTO TEÓRICO.....	18
	2.1 Aprendizado de Máquina.....	18
	2.1.1 Árvores de Decisão.....	19
	2.1.1.1 Árvores de classificação	20
	2.1.1.2 Árvores de regressão.....	21
	2.1.2 <i>Random Forest</i>	22
	2.1.3 Redes Neurais	23
	2.1.3.1 Redes Neurais <i>Feed Forward</i>	25
	2.1.3.2 Rede Neural Recorrente.....	27
	2.1.3.3 <i>Long Short-Term Memory</i>	27
3	SOLUÇÃO PROPOSTA	31
	3.1 Extração dos dados	31
	3.2 Seleção dos dados	33
	3.2.1 Predição quantitativa.....	33
	3.2.2 Predição qualitativa.....	33
	3.3 Pré-processamento dos dados	34
	3.3.1 Predição quantitativa.....	34
	3.3.2 Predição qualitativa.....	37
	3.4 Modelo de aprendizado de máquina	42
	3.4.1 Critérios de validação	43
	3.4.1.1 Predição quantitativa.....	43
	3.4.1.2 Predição qualitativa	44
4	EXPERIMENTOS E DISCUSSÕES.....	45

4.1	Desenvolvimento dos modelos	45
4.2	Experimentos	45
4.2.1	Predição quantitativa	45
4.2.1.1	Random Forest.....	46
4.2.1.2	Feed Forward em conjunto com Long Short-Term Memory	48
4.2.1.3	Comparação dos resultados	50
4.2.2	Predição qualitativa	50
4.2.2.1	Random Forest.....	53
4.2.2.2	Feed Forward.....	54
4.2.2.3	Comparação dos resultados	55
5	CONCLUSÕES	57
	REFERÊNCIAS BIBLIOGRÁFICAS	59

1 INTRODUÇÃO

1.1 Apresentação e Objeto de Pesquisa

Com a capacidade de processamento sempre crescendo, os computadores são cada vez mais capazes de gerar impacto no desenvolvimento social, uma vez que desempenham muitas tarefas em níveis inalcançáveis pelos humanos (LIMA, 2014). Influenciada diretamente por isso, uma dessas tarefas é o reconhecimento de padrões, desempenhada de maneira muito mais eficiente por técnicas de aprendizado de máquina (MASCARENHAS, 2020).

Um dos seus vários tipos, é o aprendizado supervisionado. Ele consiste em ensinar ao modelo o que responder em determinada situação por meio de entradas, as quais ilustram o problema, e saídas, as quais por sua vez indicam qual deve ser o comportamento. Além disso, dentro do aprendizado citado anteriormente, existem os modelos de regressão e de classificação, que trabalham com saídas contínuas e discretas respectivamente (RUSSEL, 2013).

De modo geral, modelos de aprendizado de máquina geram modelos preditivos, ou seja, modelos que predizem certo comportamento baseado em dados que descrevem o comportamento no passado. Sendo assim, são uma ferramenta extramente útil para as empresas. Ao encontro desse pensamento, Polat (2007) considera que somente as empresas que são capazes de prever o futuro e se adaptar às novas condições alcançam seus objetivos.

Nesse contexto, planos de assinaturas mensais oferecem serviços ou produtos mensalmente perante um pagamento automático no início de cada mês. São sinônimos de comodidade e praticidade para o cliente e previsibilidade para a empresa, por isso, representam uma tendência de modelo de negócio. Uma vez que sabe-se a quantidade de assinantes em certo momento, os meses seguintes ficam muito mais previsíveis com relação à receita para uma empresa que trabalha dessa forma do que para uma empresa que trabalha apenas com vendas avulsas (SHAABAN et al., 2012).

Este projeto visa portanto a previsão quantitativa e qualitativa dos cancelamentos de assinaturas mensais de vinho da empresa Wine. A previsão quantitativa indicará quantas assinaturas serão canceladas através de duas técnicas de aprendizado de máquina supervisionado de regressão,

para que no final seja realizada uma comparação entre elas. A previsão qualitativa irá indicar quais assinaturas mais prováveis de realizar o cancelamento no próximo mês por meio de duas técnicas de aprendizado de máquina supervisionado de classificação, também para que seja possível compará-las.

1.2 Justificativa

Com o aumento da competitividade entre as empresas, aquela que consegue se antecipar às oportunidades e problemas tem mais chance de sobreviver no mercado. Além disso, os sistemas de assinaturas tem sido grande alvo de empresas, uma vez que representam uma renda fixa, essencial para a sobrevivência e planejamento delas (SHAABAN et al., 2012).

Conseguir compreender o comportamento dos assinantes e, ainda mais, prever os cancelamentos das assinaturas permite que atitudes sejam tomadas com antecedência em vista de indicar o futuro com mais precisão. Nessa linha de pensamento, é correto considerar que o cancelamento de assinaturas é o maior problema e preocupação das grandes empresas que trabalham assim e que muitas, especialmente as de telecomunicações, estão desenvolvendo modelos de predição (AHMAD; JAFAR; ALJOUAAA, 2019). O tempo ganho em saber quem tem mais chance de cancelar a assinatura ou pelo menos quantos irão cancelar significa uma vantagem financeira enorme para as empresas uma vez que estão mantendo sua melhor fonte de renda e reter clientes custa de 5 a 10 vezes menos para as empresas do que captar novos (SHAABAN et al., 2012).

Além desse contexto e utilizando o mesmo procedimento, modelos preditivos têm aplicações em detecção de fraudes, fenômenos da natureza e são aplicados amplamente na área da saúde, com diagnósticos mais precisos e rápidos por exemplo (PEREIRA, 2017). De modo geral, representam um excelente caminho para qualquer situação que seu passado possa ser extraído em dados. Sendo assim, este projeto servirá portanto como base para aplicações e objetos de estudo com características semelhantes, por exemplo, qualquer outra empresa com sistemas de assinaturas mensais ou situações onde se necessita estudar o que o cliente sinaliza para demonstrar sua insatisfação com algum produto ou serviço.

1.3 Objetivos

1.3.1 Objetivo geral

Este trabalho tem como objetivo geral a previsão do número de cancelamentos de planos mensais de vinho da empresa Wine para o mês seguinte e também a previsão de quais assinaturas irão cancelar também no mês seguinte.

1.3.2 Objetivos Específicos

Os objetivos específicos contemplam a extração e seleção dos dados do banco de dados da Wine e o tratamento deles. Além disso, visa-se a aplicação de técnicas de regressão e classificação e, por fim, a análise dos resultados.

1.4 Estrutura do Texto

Este trabalho está distribuído da seguinte forma após a introdução já apresentada:

- **Embasamento Teórico:** nesta parte do trabalho, será estabelecida toda a explicação teórica do desenvolvimento das previsões.
- **Solução proposta:** esta seção explica como está estruturada a resolução do problema abordado por este trabalho, bem como a justificativa para cada escolha feita.
- **Experimentos e Resultados:** esta parte mostra como são compostos os experimentos que buscam solucionar o problema abordado pelo trabalho. Além disso, traz o resultado e desempenho de cada um deles.
- **Conclusões:** nesta seção será feita a comparação do resultado de todas as técnicas aplicadas nos experimentos a fim de informar qual teve melhor desempenho. E, por fim, traz quais as melhorias e próximos passos que podem ser aplicados neste projeto.

2 EMBASAMENTO TEÓRICO

2.1 Aprendizado de Máquina

Um grande foco de estudos e pesquisas tecnológicas, o aprendizado de máquina vem em constante crescimento, principalmente por suas aplicações. Ele é capaz desde identificar objetos em imagens (MONTANARI, 2015) à identificar doenças em pacientes (PEREIRA, 2017). E surgiu com a intenção de responder ao questionamento se os computadores eram capazes de pensar ou criar regras a partir de conjunto dados (TURING, 1950).

Aprendizado de máquina ou *machine learning* é a capacidade que uma máquina tem de executar uma determinada função sem ser explicitamente programada para isso (SAMUEL, 1959), ou seja, como o nome sugere, a máquina aprende a desempenhar uma função de acordo com o modelo programado. Esse aprendizado ocorre a partir de informações inseridas nela e chamadas de dados de treinamento, os quais vão fornecer as características necessárias do objeto de estudo (CARVALHO et al, 2011).

Após esse treinamento, ocorre a etapa de validação. Nesta etapa, o modelo gerado pelos dados de treinamento é testado com relação à sua fidelidade aos dados de teste. Algumas funções são utilizadas para medir essa grandeza, por exemplo funções de perda, e, de acordo com o problema, é escolhida a mais adequada (OLIVEIRA, 2016).

Além disso, modelos de aprendizado de máquina podem ter algumas classificações quanto ao método de aprendizado, neste trabalho será utilizado o aprendizado supervisionado. Este aprendizado consiste em rotular as saídas do conjunto de dados de treinamento indicando assim o caminho a ser seguido de acordo com o padrão identificado para certo dado de entrada (CARVALHO et al, 2011).

Quanto à saída do modelo, este tipo de aprendizado pode ser de regressão ou de classificação. Modelos de regressão significam que as saídas deles são contínuas, por outro lado, modelos de classificação significam que as saídas são discretas, ou seja, são números finitos que representam classificações de uma grandeza.

Modelos de classificação buscam o reconhecimento de padrões, em outras palavras, têm como objetivo classificar os objetos em grupos de acordo com suas características semelhantes identificadas (THEODORIDIS; KOUTROUMBAS, 2006). Além disso, o reconhecimento de padrões utiliza uma representação do conjunto total de dados que indique as características mais relevantes, reduzindo assim a quantidade de informação necessária para a quantidade de grupos identificados (DUDA; HART; STORK, 2012).

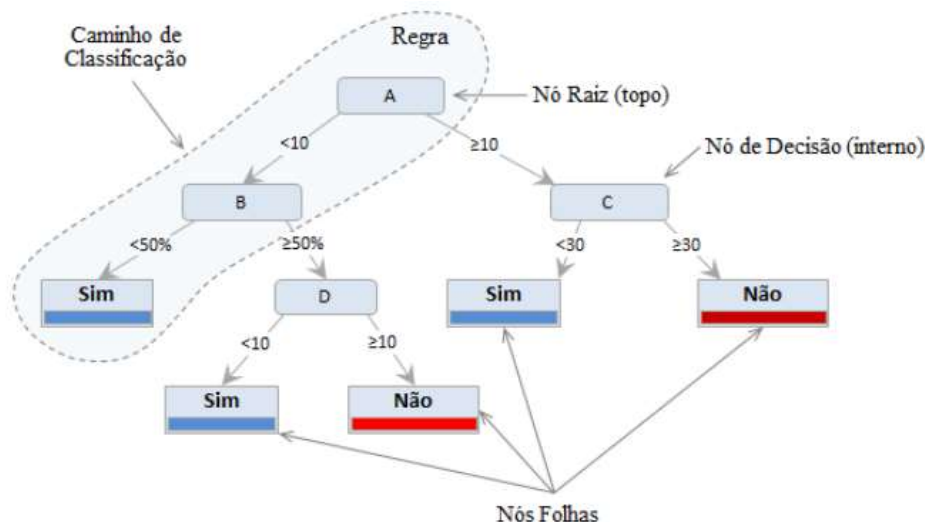
Já os modelos de regressão podem ser considerados como técnicas que visam explicar a relação entre variáveis independentes e uma variável dependente (MESQUITA; NOGUEIRA, 2015). Ao contrário dos modelos de classificação, a regressão informa como saída um número ou quantidade de certa grandeza, sendo, portanto, quantitativa.

2.1.1 Árvores de Decisão

Uma forma eficiente de aplicar um modelo de aprendizado de máquina é pela técnica de árvore de decisão. É um tipo de estrutura de dados não linear que possui um número finito de elementos ou nodos (HALMENSCHLAGER, 2002).

Essa técnica se baseia em uma solução de modelo hierárquico *top-down*, ou seja, do nó raiz em relação ao nó folha, como mostra a Figura 1. As formas de analisar os dados e decidir suas influências para construção da árvore varia de acordo com o critério utilizado, mas em todos os casos o problema é simplificado, realizando análises sobre problemas menores em subconjuntos dos dados (HALMENSCHLAGER, 2002).

Figura 1 – Estrutura da árvore de decisão



Fonte: Rodrigues (2014).

2.1.1.1 Árvores de classificação

Para classificação, existem várias técnicas para construção das árvores de decisão, neste projeto será utilizada a técnica de Gini. Desenvolvido por Breiman, esta técnica visa medir o grau de impureza de cada característica do conjunto de dados de treinamento em relação às suas saídas (ALSABTI, 1998).

Para preenchimento dos nós, é medido o grau de impureza para cada característica em relação às classes, o qual é descrito pela seguinte equação:

$$Gini(S) = 1 - \sum_{i=0}^k p_i^2 \quad (2.1)$$

onde S é o conjunto de dados, p_i é a probabilidade relativa à classe i em S e k é o número de classes.

Sendo assim, a característica que ocupa o nó é a característica que possui menor grau de impureza. Por fim, esse processo se repete para os nós abaixo e encerra quando a impureza

de um nó é menor do que se for usada uma nova característica para separá-lo, então ele se torna um nó folha e a árvore termina naquele caminho (ALSABTI, 1998).

2.1.1.2 Árvores de regressão

Para a árvore de regressão, embora também existam outras técnicas de construção da árvore, a utilizada neste trabalho será a de redução da variância de um conjunto S , denominada de V^2 (BREIMAN, 1984). Assim como a técnica utilizada para árvores de classificação, esta técnica divide o problema em situações menores, observando individualmente cada característica do conjunto de dados de treinamento e analisando sua influência na informação de saída.

Para construção da árvore, inicia-se o processo de medição da variância no nó raiz pela seguinte equação:

$$V = \frac{\sum_{i=0}^n (y_i - y_m)^2}{n} \quad (2.2)$$

onde n é o número de elementos no conjunto de dados de treinamento, y_i é o valor da saída observada no conjunto de dados de treinamento do i -ésimo elemento e y_m é o valor médio da saída observada no conjunto de dados de treinamento, obtido pela equação a seguir:

$$y_m = \frac{\sum_{i=0}^n (y_i)}{n} \quad (2.3)$$

Sendo assim, a característica que ocupa o nó é a característica que possui menor valor de variância. Por fim, esse processo se repete para os nós abaixo e encerra quando a variância de um nó é menor do que se for usada uma nova característica para separá-lo, então ele se torna um nó folha e a árvore termina naquele caminho (BREIMAN, 1984).

2.1.2 *Random Forest*

A estimativa de parâmetros a partir de um único modelo sobre o conjunto de dados de treinamento envolve algumas escolhas subjetivas e isso muitas vezes torna o resultado subestimado ou super otimista e, em muitos casos, pouco flexível com relação às inúmeras correlações entre as informações contidas nele (AUGUSTIN, 1997). Como um caminho para mitigar esses efeitos, o método de *bootstrap* consiste em aumentar o número de modelos gerados e ponderar seus resultados, trazendo assim mais informações sobre as relações entre os subconjuntos do conjunto de dados de treinamento (EFRON, 1993).

Nesse contexto, com o intuito de aumentar a estabilidade e a percepção sobre as relações entre as características de entrada, o *bagging* (*bootstrap aggregating*) consiste em gerar alguns modelos com os subconjuntos gerados pelo reamostragem *bootstrap*, e, a partir delas, obter modelos preditivos para estimativas (BREIMAN, 1996). Com essas técnicas aplicadas, Breiman (1996) e Optiz e Maclin (1999) afirmam que os erros de predição tendem a ser menores do que quando é aplicado apenas um modelo para estimar saídas, afirmação essa que foi confirmada por Bulhuman e Yü (2002) quando pontuaram que aplicar a técnica de *bagging* reduz valores de funções de perda como erro quadrático médio.

Sendo assim, desenvolvida por Leo Breiman e Adele Cutler, a técnica de *Random Forest* se trata da aplicação da técnica de *bagging* com modelos preditivos de árvore de decisão, seja para modelos preditivos de regressão ou classificação (BREIMAN, 2001).

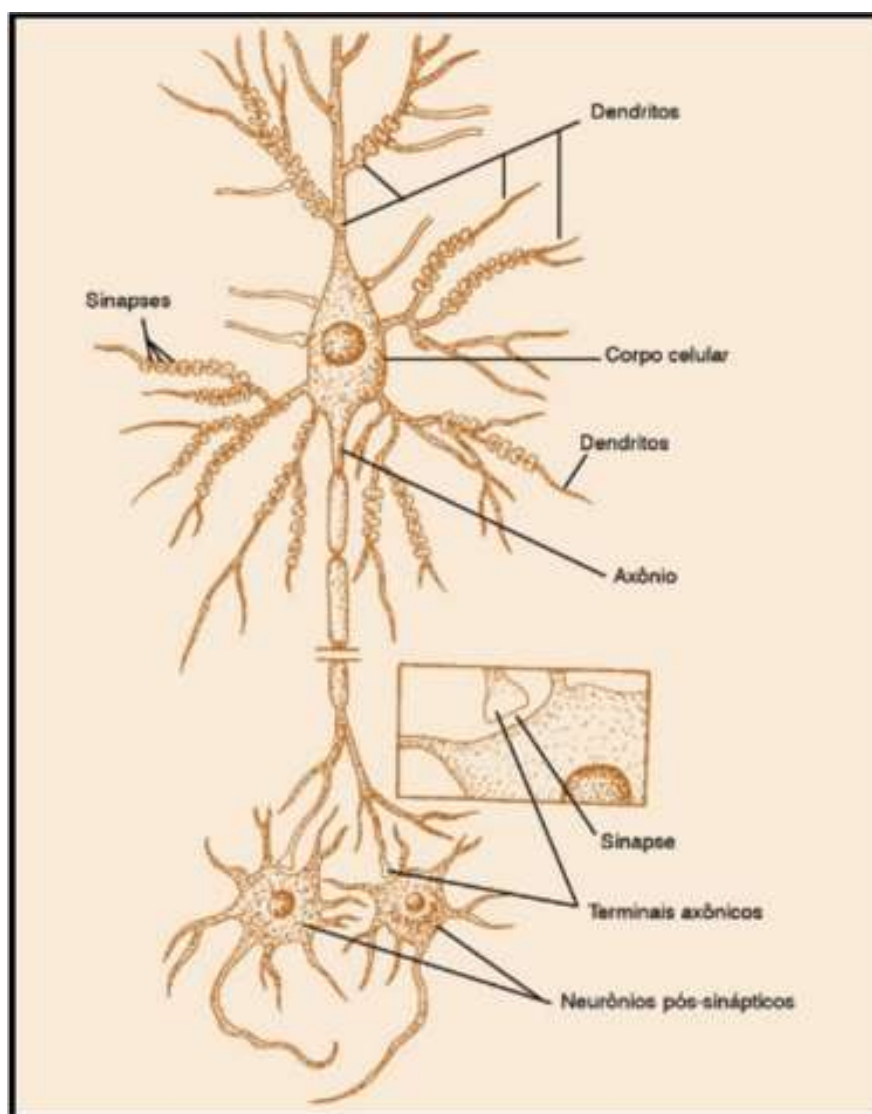
A aplicação do método de *Random Forests* executa a seguinte sequência (DANTAS; DONADIA, 2013):

1. Obter j amostras do conjunto de dados de treinamento.
2. Aplicar técnicas de árvore de decisão de regressão ou classificação sobre cada amostra.
3. Estimar valores de saída a partir de médias ponderadas sobre as estimativas realizadas pelas árvores de decisão geradas no passo anterior.

2.1.3 Redes Neurais

As redes neurais, como mostra a Figura 2, são sistemas computacionais que imitam o funcionamento do cérebro humano para aprender e desempenhar algumas tarefas humanas com mais precisão e com maior quantidade de informações, como por exemplo reconhecimento de padrões (KELLER; PRIDDY, 2005). Segundo Haykin (2008), uma rede neural é um processador paralelo massivamente distribuído que possui uma propensão natural à armazenar conhecimento adquirido pela experiência e torná-los disponível para uso.

Figura 2 - Esquema de ligação entre os neurônios no cérebro humano.



Fonte: Izquierdo (2011).

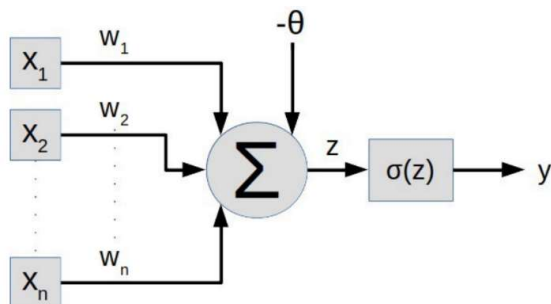
Elas são compostas por uma camada de entrada, onde recebe os dados, uma camada de saída, onde ela fornece uma resposta para a entrada e uma camada ou várias camadas intermediárias, sendo chamada de rede neural quando existe apenas uma camada intermediária e de rede neural profunda quando existem mais de uma camada intermediária (HAYKIN, 2008). Além disso, as camadas por sua vez são compostas por neurônios, eles são funções não lineares que transformam variáveis de entrada em variáveis de saída. Os sinais de entrada são multiplicados por pesos que agem de forma semelhante à intensidade sináptica do neurônio biológico e, posteriormente, somados (PONTES, 2011), como mostra a próxima equação:

$$z = x_0w_0 + x_1w_1 + x_2w_2 + \dots x_mw_m - \theta \quad (2.4)$$

onde z representa o valor total recebido pelo neurônio, x as entradas, m o número de neurônios na camada anterior, w os pesos que multiplicam cada entrada e θ é o coeficiente linear.

Após isso, como ilustra a Figura 3, passam pela função de ativação, elas têm, como principal função, controlar o valor de saída, para evitar que ele seja muito elevado. Os primeiros estudos voltados à elas foram descritos por Elliott (1993), quando estudou os impactos sobre redes neurais.

Figura 3 - Esquema de funcionamento de um neurônio



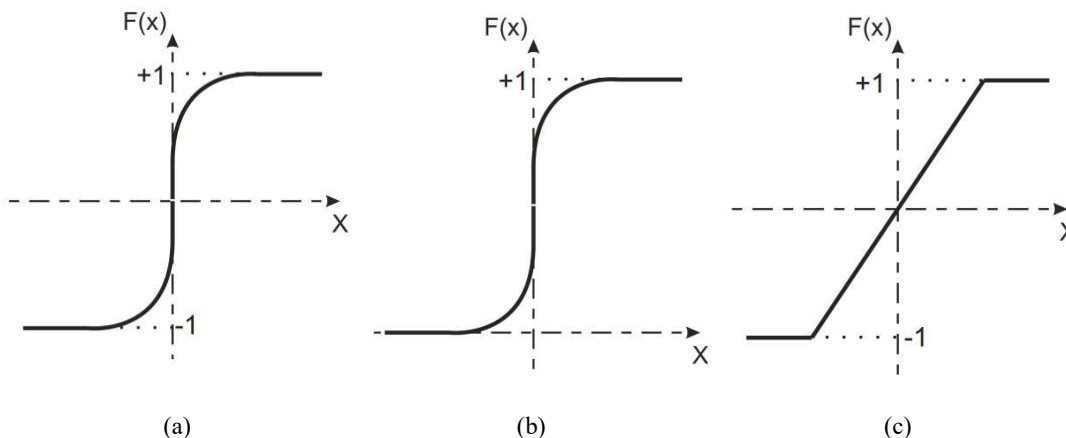
Fonte: Scalabrin (2019).

onde σ é a função de ativação e y a saída do neurônio.

Essas funções utilizam os valores recebidos das camadas anteriores para calcular as médias ponderadas com os pesos, também conhecidos como *bias* (SANTOS, 2019). Além disso,

podem ser lineares ou não lineares e umas das mais utilizadas são a função tangente hiperbólica, sigmóid e rampa, como mostra o Gráfico 1.

Gráfico 1 – Funções de ativação: (a) função tangente hiperbólica, (b) função sigmóid e (c) função rampa



Fonte: (a), (b) e (c) (SANTOS, 2008).

2.1.3.1 Redes Neurais *Feed Forward*

Todo o aprendizado que ocorre em redes neurais é originado pela correção da diferença entre o valor estimado na saída e o valor estabelecido para as entradas em questão, ou seja, é originado da correção dos valores dos pesos que alimentam os neurônios para que o erro fique cada vez menor. Este processo se dá por meio de dois movimentos e conhecido como processo de retropropagação (SILVA, 1998).

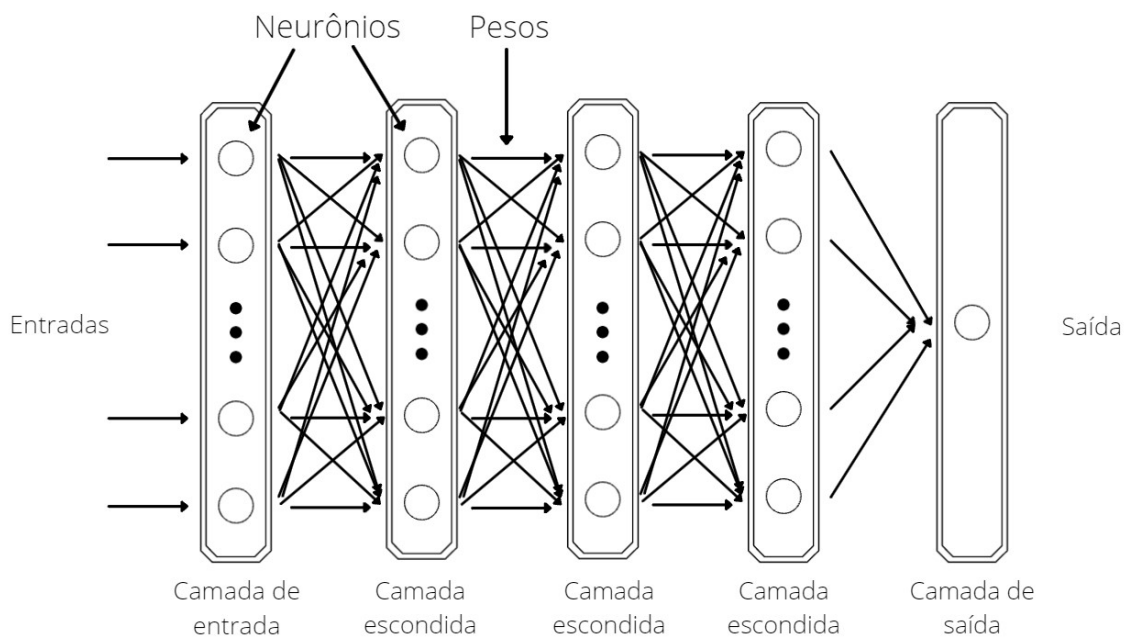
O primeiro movimento, conhecido como *feed forward* se dá no sentido da entrada à saída da rede neural. Os valores são recebidos pela camada de entrada e se propagam pela rede multiplicados pelos pesos e acionando as funções de ativação que compõem os neurônios das camadas posteriores, originando assim valores de saída (SILVA, 1998).

Essas estimativas geradas são comparadas com os valores esperados para a entrada, caso exista uma diferença entre eles, um movimento no sentido contrário ocorre. Esse movimento tem a função de ajustar os valores dos pesos de entrada dos neurônios e ficou conhecido como *backpropagation*. Foi introduzido por Hinton, Rumelhart e Willians (1998) e é a forma pela

qual essa técnica de aprendizado de máquina extrai conhecimento e se molda às características de cada situação.

Sendo assim, redes neurais *feed forward* são os tipos mais simples de redes neurais, as quais só possuem os movimentos de *feed forward* e de *backpropagation*, como ilustra a Figura 4. Sendo portanto acíclicas (SANTOS, 2019).

Figura 4 - Rede neural *Feed Forward*



Fonte: Próprio Autor.

2.1.3.2 Rede Neural Recorrente

Um tipo mais robusto de rede neural são as RNN (*Recurrent Neural Network*), elas, diferentemente das redes neurais *Feed Forward*, não possuem somente um sentido de propagação, as camadas ocultas recebem informações tanto das camadas de entrada quanto da própria relativo às iterações passadas (LECUN; BENGIO; HINTON, 2015). Na prática, isso funciona como uma memória, ou seja, a saída atual sofre influência tanto da entrada atual, quanto de entradas passadas em um certo período de tempo, como mostram as seguintes equações:

$$h_t = \Phi(b_h + xw_h + h_{t-1}w_h) \quad (2.5)$$

$$SN_t = b_0 + h_t w_0 \quad (2.6)$$

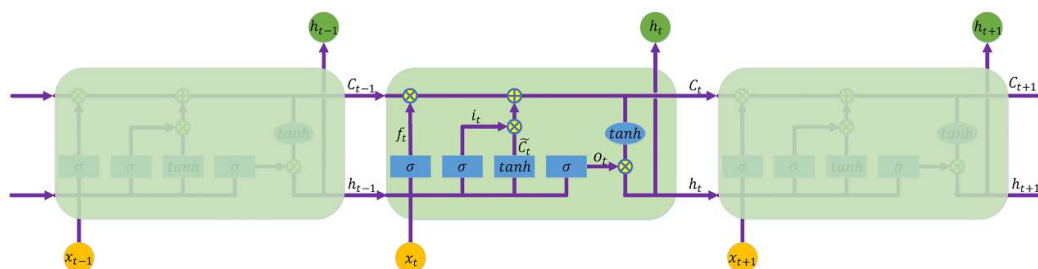
onde SN representa a saída do neurônio, x a entrada, Φ é a função sigmóide e h carrega as informações do estado anterior.

Na teoria, RNN são capazes de lidar com todos os problemas de dependência às entradas passadas, entretanto, na prática, alguns deles não são alcançados por elas, principalmente aqueles com dependência temporal a longo prazo, como explicam LeCun, Bengio e Hinton (2015).

2.1.3.3 Long Short-Term Memory

Explicitamente projetadas para resolver o tipo de problema que surgiu na seção 2.1.3.2, existem as redes neurais LSTM (*Long Short-Term Memory*), as quais foram introduzidas por Hochreiter e Schmidhuber (1997). Enquanto as RNN possuem apenas uma camada de repetição, elas possuem um sistema mais complexo com quatro mecanismos diferentes, como ilustra a Figura 5.

Figura 5 – LSTM.



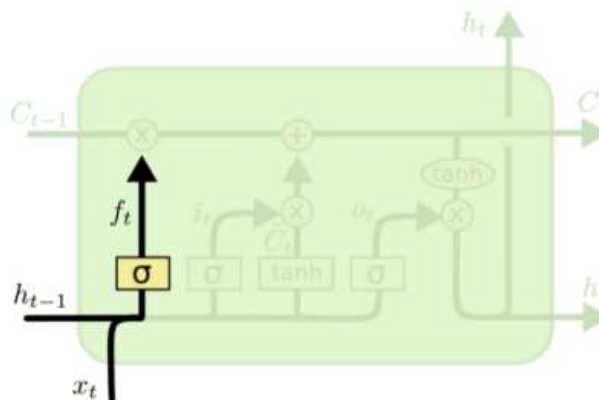
Fonte: Bao, Yue e Rao (2017).

Num primeiro momento, a célula da LSTM decide qual informação manter e qual descartar da informação de entrada atual, como mostra a Figura 6, onde os valores de f_t podem variar entre 0 e 1, significando esquecer ou manter respectivamente, a seguinte equação define f_t :

$$f_t = \Phi(w_f[h_{t-1}, x_t] + b_f) \quad (2.7)$$

onde x é a entrada atual, Φ é a função sigmóide, h carrega as informações do estado anterior e w é o peso aplicado à entrada.

Figura 6 – Mecanismo de esquecimento da LSTM.



Fonte: Olah (2015).

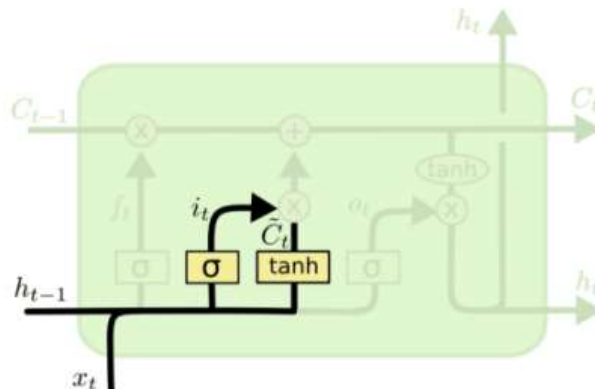
Em seguida, a célula decide qual informação armazenar em seu estado atual, como mostra a Figura 7, e as seguintes equações encontram os valores de i_t e C_t :

$$i_t = \Phi(w_i[h_{t-1}, x_t] + b_i) \quad (2.8)$$

$$C_t = \tanh(w_c[h_{t-1}, x_t] + b_c) \quad (2.9)$$

onde \tanh é a tangente hiperbólica e C é o estado do neurônio.

Figura 7 – Mecanismo de armazenamento da LSTM

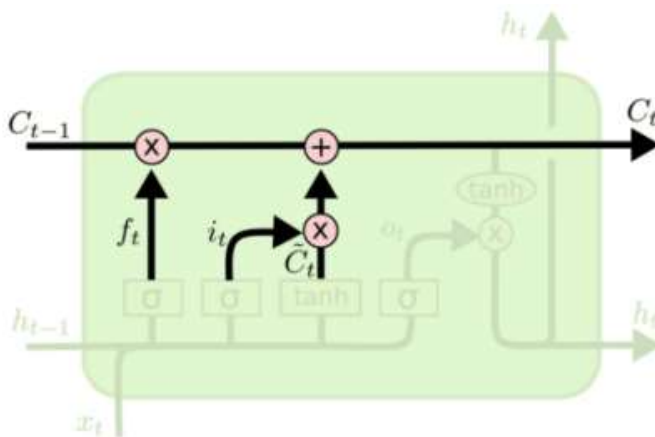


Fonte: Olah (2015).

Com todas essas informações processadas e escolhidas, a célula coloca em prática a seleção e escolhe o que deve ficar ou não nos estados antigo e atual, como mostra a Figura 8, uma outra fórmula para o valor de C_t é mostrada a seguir:

$$C_t = f_t C_{t-1} + i_t C_t \quad (2.10)$$

Figura 8 – Mecanismo de atualização do neurônio da LSTM.



Fonte: Olah (2015).

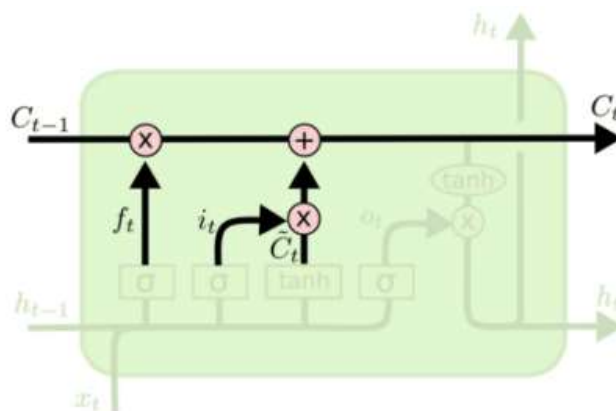
Passadas as etapas de processamento de estado de célula, ela, por fim, determina qual sua saída, como mostra a Figura 9, e as equações para os valores de O_t e h_t são mostradas na sequência:4

$$O_t = \Phi(w_o[h_{t-1}, x_t] + b_o) \quad (2.11)$$

$$h_t = O_t \tanh(C_t) \quad (2.12)$$

onde O é a função de saída do neurônio.

Figura 9 – Mecanismo de saída do neurônio da LSTM



Fonte: Olah (2015).

3 SOLUÇÃO PROPOSTA

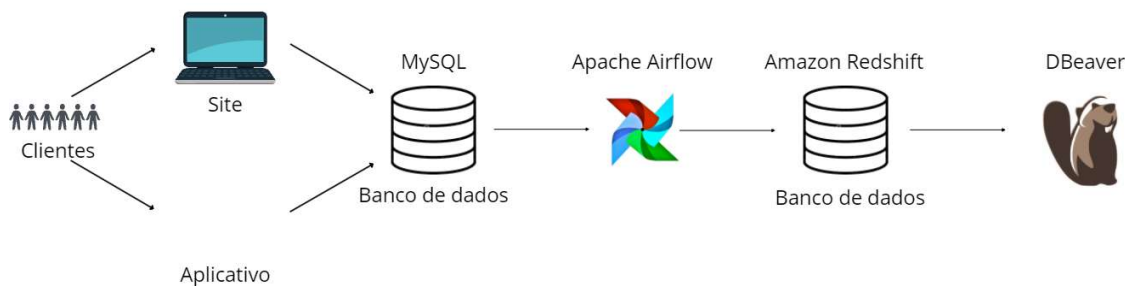
3.1 Extração dos dados

Este projeto visa estudar a previsão do cancelamento de assinaturas de vinho analisando o comportamento dos clientes, sendo assim, todas as informações utilizadas são provenientes de interações deles com a empresa. Embora existam algumas por meio físico, neste estudo serão coletadas as que ocorrem por meio virtual, mais especificamente por aplicações *Web e Mobile*, ou seja, movimentos dos clientes pelo site da empresa ou pelo aplicativo respectivamente.

Dentro dessas aplicações, os clientes podem realizar procedimentos de busca e compra de produtos, cancelamento de compras, assinatura e cancelamento de planos, cadastro de informações pessoais, indicação de assinatura de clube para amigos, cadastro e escolha de informações referentes à compras, como endereço de entrega e modo de pagamento, gerenciamento de conta sobre a política de *cashback* entre outros.

Todos esses procedimentos geram dados que abastecem um banco de dados MySQL por meio da linguagem SQL, como ilustra a Figura 10. O MySQL, por sua vez, é um banco relacional, ou seja, os dados são armazenados nele por meio de tabelas. Bancos de dados relacionais de um modo geral utilizam chaves ou identificadores para cada unidade existente, isso permite relacionar informações entre tabelas e também localizar informações de maneira eficiente. Além disso, são otimizados para armazenar linhas de dados e, por essa razão, são utilizados geralmente para aplicações transacionais como neste caso. Entretanto, por ser essencial para os processos da empresa, participar ativamente das vendas, pagamentos e entregas por exemplo, ou seja, sofrer alterações em tempo real com as interações dos clientes, as consultas aos dados não ocorrem nele, justamente para otimizar tanto os processos como as consultas.

Figura 10 - Extração dos dados



Fonte: Próprio Autor.

Para realização de consultas, os dados são armazenados em um segundo banco de dados, o Amazon Redshift. Se trata de um banco colunar, ou seja, os dados são armazenados em colunas e são desenvolvidos para recuperação rápida delas. Sendo assim, muito mais adequado para aplicativos analíticos. Toda essa movimentação de dados do banco MySQL para o banco Redshift, ou seja, do banco de dados relacional para o banco colunar, é orquestrada pelo *software* AirFlow (Figura 10). Neles são desenvolvidos códigos na linguagem Python (ROSSUM, 1995), através da qual, com a utilização de bibliotecas, realiza a extração, transformação e inserção dos dados por meio da linguagem SQL, também conhecidos como processos de ETL's.

ETL's são definidas por Pereira (2000) como ferramentas que têm como objetivo extrair, transformar e realizar a carga de dados em uma *Data Warehouse*. Esta por sua vez, é geralmente o ponto central de dados de uma empresa, foram projetados para dar suporte para a inteligência do negócio. Souza (2003) complementa que esse conceito foi introduzido para posicionar as empresas mais estrategicamente, tornando-as mais competitivas, maximizando lucros e reduzindo índices de erro. Por fim, para consultar e extrair os dados do banco colunar Amazon Redshift é utilizada uma aplicação analítica com linguagem SQL, neste caso o *software* DBeaver (Figura 10).

3.2 Seleção dos dados

Bancos de dados reais possuem muitas vezes um número muito elevado de registros. Por outro lado, modelos de aprendizado de máquina tem limitações quanto ao processamento de informações. Esta etapa portanto consiste na seleção de um subconjunto de dados que represente da melhor forma o conjunto total de dados que ilustra a situação estudada (BEZERRA, 2017).

3.2.1 Predição quantitativa

Todas as empresas com o passar do tempo modificam seus produtos para se adequar as necessidades delas e dos clientes. Sendo assim, no passado, foram realizadas mudanças quanto às assinaturas de vinho na empresa, entre elas estão a criação e exclusão de modalidades e a criação de cupons de desconto para certos clientes. Portanto, visando adequar ao máximo o resultado desta parte do projeto à realidade da empresa no momento, serão selecionados apenas dados de Agosto de 2017 até hoje.

3.2.2 Predição qualitativa

Com a mesma justificativa da seção 3.2.1, para essa amostra, serão selecionados apenas dados com registro a partir do ano de 2019, uma vez que este projeto visa encontrar o comportamento dos clientes e é mais coerente analisar este comportamento o mais próximo da data atual possível, uma vez que hábitos e processos da empresa mudam constantemente.

Outro critério de seleção será em função da proporção de dados onde a assinatura foi cancelada no mês ou não. De acordo com Prati (2006), quando as classes em conjuntos de dados de treinamento estão muito desbalanceadas, os modelos têm dificuldade de encontrar as características das classes que estão em menor quantidade. Aspecto esse confirmado por Braga e Castro (2011), quando explicam que modelos com classes desbalanceadas tendem a sinalizar as categorias que estavam em maior número no conjunto de dados de treinamento, resultando assim em uma baixa taxa de reconhecimento das classes minoritárias. Assim, os dados terão a mesma quantidade tanto de registros de assinaturas canceladas no próximo mês como de registros de assinaturas que não foram canceladas.

3.3 Pré-processamento dos dados

De acordo com García, Herrera e Luengo (2015), os dados provenientes dos bancos de dados reais frequentemente não estão adequados para a utilização em modelos de otimização. Além disso, como pontua Bezerra (2017), o pré-processamento de dados é uma etapa essencial para a mineração que contribui significativamente para o conhecimento do conjunto de dados que serão utilizados.

Sendo assim, esta etapa consiste em realizar, por exemplo, a limpeza, integração, transformação e normalização dos dados, variando a execução de cada ponto citado de acordo com o objetivo do modelo. É portanto a preparação dos dados para que eles atendam aos requisitos dos algoritmos de otimização (BEZERRA, 2017).

3.3.1 Predição quantitativa

Como pontua Silva (2019), modelos de aprendizado de máquina necessitam trabalhar com valores numéricos, uma vez que não podem processar diretamente informações no formato textual. Para isso, será utilizada a categorização dos dados que estão em formato textual, ou seja, cada registro textual diferente será representado por uma categoria ou classe, através de um número.

Em bancos de dados reais, é comum que registros tenham inconsistências ou ruídos decorrentes, por exemplo, de mudança de processos da empresa. Assim, Han e Kamber (2000) indicam que as tarefas de limpeza de dados têm como objetivo minimizar os efeitos dos ruídos nos conjuntos de dados, e, para isso, identificam inconsistências e modificam os conteúdos das informações que as geram.

Sendo assim, técnicas de limpeza de dados foram aplicadas para corrigir algumas informações registradas no banco de dados que possuíam nomes diferentes, entretanto têm o mesmo significado. Como exemplo disso, os planos de assinaturas possuíam um determinado nome e após algum tempo foram modificados por razões internas.

Outro aspecto a ser considerado são os *outliers*, segundo Knorr (2002) são objetos que não estão de acordo com o comportamento geral ou modelo de dados. São objetos completamente diferentes ou inconsistentes com o restante dos dados. Além disso, os caracteriza como valores extremos, um dado legítimo mas inesperado, um contaminante ou valor que foi medido ou gravado incorretamente.

Com isso, existem várias formas de identificá-los, neste trabalho serão empregadas técnicas baseadas em modelos matemáticos. Elas assumem que os dados seguem a distribuição normal e utilizam as métricas de média e desvio-padrão (TANURE, 2010). Também para este modelo, será aplicado o agrupamento dos dados, isso consistirá na soma dos registros no banco de dados com informações semelhantes às informações de entrada do modelo.

Além disso, a normalização dos dados consiste na mudança de escala dos números, convertendo eles para que, proporcionalmente aos seus valores originais, fiquem entre os valores 0 e 1 ou -1 e 1, de acordo com o objetivo desejado (HAN, 2000), entretanto deve-se ter cuidado para que esse procedimento não altere a forma como os dados estão distribuídos (BEZERRA, 2017). Portanto, como explicam Keller e Priddy (2005), para modelos de redes neurais, é mais adequado que se utilize valores normalizados, pois eles trazem benefícios como redução no tempo de treinamento, melhor acurácia e evita a saturação dos neurônios. E para os modelos gerados a partir de redes neurais, será aplicada a normalização, explicada pela seguinte equação:

$$U = \frac{d}{d_m} \quad (3.1)$$

onde U é o valor normalizado, d é o valor original e d_m é máximo valor encontrado no conjunto de dados.

O conjunto de dados final será composto por 15 grandezas de entrada e 1 de saída informando o que se deseja prever nesta parte do trabalho, a quantidade de cancelamentos mensais. A tabela informa as grandezas de cada coluna e o conceito sobre cada informação está explicado a seguir no Quadro 1:

Quadro 1 - Atributos do conjunto de dados da predição quantitativa

Atributo	Tipo	Descrição do atributo
Mês	Numérico	Valor numérico do mês no ano.
Assinaturas Essenciais	Numérico	Quantidade de assinaturas ativas da modalidade essenciais
Assinaturas Surpreendentes	Numérico	Quantidade de assinaturas ativas da modalidade surpreendentes
Assinaturas Notáveis	Numérico	Quantidade de assinaturas ativas da modalidade notáveis
Assinaturas Singulares	Numérico	Quantidade de assinaturas ativas da modalidade singulares
Assinaturas Refrescantes	Numérico	Quantidade de assinaturas ativas da modalidade refrescantes
Assinaturas Espumantes	Numérico	Quantidade de assinaturas ativas da modalidade espumantes
Assinaturas	Numérico	Quantidade de assinaturas ativas naquele mês
Cupom expirando	Numérico	Quantidade de assinaturas ativas com cupom de desconto encerrando naquele mês
Migração nos últimos 30 dias	Numérico	Quantidade de assinaturas que mudaram de modalidade nos últimos 30 dias

Migração entre 30 e 60 dias	Numérico	Quantidade de assinaturas que mudaram de modalidade entre 30 e 60 dias atrás
Migração entre 60 e 90 dias	Numérico	Quantidade de assinaturas que mudaram de modalidade entre 60 e 90 dias atrás
Alteração de endereço nos últimos 30 dias	Numérico	Quantidade de assinaturas que mudaram de endereço de entrega nos últimos 30 dias
Alteração de endereço entre 30 e 60 dias	Numérico	Quantidade de assinaturas que mudaram de endereço de entrega entre 30 e 60 dias atrás
Alteração de endereço entre 60 e 90 dias	Numérico	Quantidade de assinaturas que mudaram de modalidade entre 60 e 90 dias atrás
Cancelamentos	Numérico	Quantidade de cancelamentos

Fonte: Próprio Autor.

3.3.2 Predição qualitativa

Para o conjunto de dados que irá gerar este modelo, será aplicada a categorização e limpeza dos dados, assim como no modelo de predição qualitativa e, embora muitos dados tenham características diferentes, as mesmas razões se aplicam. Por fim, para as redes neurais, será aplicado a normalização, também como explicada na seção 3.3.1. E obedeceram à estrutura do Quadro 2:

Quadro 2 - Atributos do conjunto de dados da predição qualitativa

Atributo	Tipo	Descrição do atributo
Mês	Numérico	Valor numérico do mês no ano.
Dias de assinatura	Numérico	Há quantos dias a assinatura está ativa

Modalidade	Numérico	Número de 1 à 7 representando as modalidades Essenciais, Surpreendentes, Notáveis, Singulares, Refrescantes, Espumantes, Winebox do Seu Jeito e Outras respectivamente.
Migração	Binário	A assinatura já migrou de plano ou não
Migração 1 mês atrás	Binário	A assinatura migrou de plano nos últimos 30 dias
Migração 2 meses atrás	Binário	A assinatura migrou de plano entre os últimos 30 dias e 60 dias
Migração 3 meses atrás	Binário	A assinatura migrou de plano entre os últimos 60 dias e 90 dias
Mudança de clube	Binário	A assinatura já mudou de modalidade ou não
Mudança de clube 1 mês atrás	Binário	A assinatura mudou de clube nos últimos 30 dias
Mudança de clube 2 meses atrás	Binário	A assinatura mudou de clube entre os últimos 30 dias e 60 dias
Mudança de clube 3 meses atrás	Binário	A assinatura mudou de clube entre os últimos 60 dias e 90 dias
Mudança de endereço	Binário	A assinatura já mudou de endereço ou não
Mudança de endereço 1 mês atrás	Binário	A assinatura mudou de endereço nos últimos 30 dias
Mudança de endereço 2 meses atrás	Binário	A assinatura mudou de endereço entre os últimos 30 dias e 60 dias
Mudança de endereço 3 meses atrás	Binário	A assinatura mudou de endereço entre os últimos 60 dias e 90 dias
Total de compras	Numérico	Quantidade de compras realizadas pelo cliente

Média de compras	Numérico	Média mensal de compras realizadas pelo cliente
Compras 1 mês atrás	Numérico	Quantidade de compras realizadas pelo cliente nos últimos 30 dias
Compras 2 meses atrás	Numérico	Quantidade de compras realizadas pelo cliente entre os últimos 30 dias e 60 dias
Compras 3 meses atrás	Numérico	Quantidade de compras realizadas pelo cliente entre os últimos 60 dias e 90 dias
Crescimento de compras 1 mês atrás	Binário	Houve crescimento ou não nas compras dos últimos 30 dias em relação às compras entre os últimos 30 e 60 dias
Crescimento de compras 1 mês atrás	Binário	Houve crescimento ou não nas compras entre os últimos 30 e 60 dias em relação às compras entre os últimos 60 e 90 dias
Compras acima da média 1 mês atrás	Binário	Quantidade de compras realizadas pelo cliente acima da média mensal nos últimos 30 dias
Compras acima da média 2 meses atrás	Binário	Quantidade de compras realizadas pelo cliente acima da média mensal entre os últimos 30 dias e 60 dias
Compras acima da média 3 meses atrás	Binário	Quantidade de compras realizadas pelo cliente acima da média mensal entre os últimos 60 dias e 90 dias
Valor total de compras	Numérico	Valor total em compras de produtos
<i>Ticket</i> médio	Numérico	Valor médio das compras já realizadas
Valor médio mensal	Numérico	Valor médio em compras por mês
Valor de compras 1 mês atrás	Numérico	Valor da soma das compras realizadas nos últimos 30 dias
Valor de compras 2 meses atrás	Numérico	Valor da soma das compras realizadas entre os últimos 30 e 60 dias

Valor de compras 3 meses atrás	Numérico	Valor da soma das compras realizadas entre os últimos 60 e 90 dias
Crescimento de valor 1 mês atrás	Binário	Houve crescimento ou não no valor em compras dos últimos 30 dias em relação ao valor em compras entre os últimos 30 e 60 dias
Crescimento de valor 2 meses atrás	Binário	Houve crescimento ou não no valor em compras entre os últimos 30 e 60 dias em relação ao valor em compras entre os últimos 60 e 90 dias
Valor acima da média 1 mês atrás	Binário	Valor em compras realizadas pelo cliente acima da média mensal nos últimos 30 dias
Valor acima da média 2 meses atrás	Binário	Valor em compras realizadas pelo cliente acima da média mensal entre os últimos 30 e 60 dias
Valor acima da média 3 meses atrás	Binário	Valor em compras realizadas pelo cliente acima da média mensal entre os últimos 60 e 90 dias
<i>Ticket</i> médio 1 mês atrás	Numérico	Valor médio das compras realizadas nos últimos 30 dias
<i>Ticket</i> médio 2 meses atrás	Numérico	Valor médio das compras realizadas entre os últimos 30 e 60 dias
<i>Ticket</i> médio 3 meses atrás	Numérico	Valor médio das compras realizadas entre os últimos 60 e 90 dias
Crescimento de <i>ticket</i> médio 1 mês atrás	Binário	Houve crescimento ou não no valor médio em compras dos últimos 30 dias em relação ao valor médio em compras entre os últimos 30 e 60 dias

Crescimento de <i>ticket</i> médio 2 meses atrás	Binário	Houve crescimento ou não no valor médio em compras entre os últimos 30 e 60 dias em relação ao valor médio em compras entre os últimos 60 e 90 dias
<i>Ticket</i> médio acima da média 1 mês atrás	Binário	Valor médio em compras realizadas pelo cliente acima da média mensal nos últimos 30 dias
<i>Ticket</i> médio acima da média 2 meses atrás	Binário	Valor médio em compras realizadas pelo cliente acima da média mensal entre os últimos 30 e 60 dias
<i>Ticket</i> médio acima da média 3 meses atrás	Binário	Valor médio em compras realizadas pelo cliente acima da média mensal entre os últimos 60 e 90 dias
Já comprou	Binário	Se já houve compra de algum produto realizada pelo cliente
Dias última compra	Numérico	Há quantos dias foi realizada a última compra
Valor última compra	Numérico	Valor da última compra realizada
Cancelamentos	Numérico	Quantidade de assinaturas canceladas do cliente
Expurgos	Numérico	Quantidade de assinaturas expurgadas do cliente
Pulos	Numérico	Quantidade de mensalidades puladas
Indicado	Binário	Se assinou por indicação

Quantidade indicações	Numérico	Quantas assinaturas já indicou
Dias última indicação	Numérico	Há quantos dias realizou a última indicação
Última indicação 1 mês	Binário	Se a última indicação ocorreu nos últimos 30 dias
Última indicação 2 meses	Binário	Se a última indicação ocorreu entre os últimos 30 e 60 dias
Última indicação 3 meses	Binário	Se a última indicação ocorreu entre os últimos 60 e 90 dias
Cancelamento	Binário	Se realizou o cancelamento desta assinatura no mês

Fonte: Próprio Autor.

3.4 Modelo de aprendizado de máquina

Neste trabalho, os principais critérios avaliados para a escolha das técnicas de aprendizado de máquina para os modelos serão (HAN, 2000):

1. Desempenho: Representa o processamento necessário para que o modelo seja gerado.
2. Escalabilidade: Representa a capacidade que a técnica tem de gerar modelos a partir de grandes quantidades de dados.
3. Robustez: Representa adaptabilidade da técnica com relação aos *outliers* presentes no conjunto de dados de treinamento, gerando assim classificações corretas mesmo na presença delas.
4. Interpretabilidade: Representa o quanto o modelo é compreensível para outro usuário que não participou do seu desenvolvimento.

3.4.1 Critérios de validação

De acordo com Bezerra (2017), a acurácia preditiva permite a validação do modelo de classificação gerado, bem como a comparação do desempenho entre diferentes modelos gerados sobre uma situação estudada, isso acontece porque ela informa a capacidade que o modelo tem de prever corretamente as classes de objetos desconhecidos.

O aspecto citado anteriormente diz respeito à modelos de classificação, entretanto se aplica também a modelos de regressão, como pontua Oliveira (2016) ao dizer que o objetivo da avaliação dos modelos é estimar a probabilidade de um modelo realizar previsões corretas quando um novo dado é apresentado. Além disso, complementa que é fundamental poder estimar corretamente essa capacidade do modelo para medir a confiabilidade em problemas reais.

3.4.1.1 Predição quantitativa

Sendo assim, os modelos de regressão neste trabalho utilizarão como métrica de validação a validação-cruzada ou o *cross-validation*. Métrica essa que utiliza técnicas de estatística para estimar a confiabilidade do modelo (DUDA; HART; STORK, 2001).

Para isso, serão utilizados dois subconjuntos de dados, o de treinamento e o de teste. O de treinamento será utilizado pelo modelo para extrair as características dos dados e identificar seus padrões (BEZERRA, 2017), onde também é responsável por estabelecer as correlações entre as informações de entrada e saída, bem como entre os dados de entrada.

Os dados de teste por sua vez são o subconjunto que será utilizado para testar o modelo, permitindo assim a validação dele, ele representa uma simulação do mundo real e conseqüentemente do desempenho do modelo em situações reais e fora do conjunto de dados selecionados (BEZERRA, 2017).

A técnica de validação cruzada utilizada neste trabalho será o *hold-out*, ela consiste em separar o conjunto de dados em dois, como explicado anteriormente, normalmente 70% dos dados para treinamento e 30% para teste (HAN; KAMBER, 2000). Além disso, é comum que esse

procedimento seja repetido k vezes e a acurácia final seja calculada pela média aritmética dos valores de acurácia de cada dupla de treinamento e teste, dessa forma, visa-se diminuir o efeito de casos isolados e trazer mais estabilidade à medida (BEZERRA, 2017). Por fim, será utilizada a média percentual do erro absoluto (MAPE), alcançada pela seguinte equação:

$$MAPE = \frac{\sum_{i=0}^n |v_{ei} - v_{ri}|}{n} \times 100 \quad (3.2)$$

onde v_{ei} é o i -ésimo valor estimado, v_{ri} é o valor real e n é o número de amostras no conjunto de dados de teste.

3.4.1.2 .Predição qualitativa

Para a predição qualitativa será utilizada também a técnica *hold-out*. Entretanto, para isso, será aplicada a matriz de confusão. Uma matriz de confusão consiste na frequência de erros e acertos de um modelo de classificação colocados em uma tabela (VASCONCELOS, 2017). Além disso, ela fornece um detalhamento de desempenho do modelo, pois indica o número de classificações corretas comparado ao conjunto de dados de teste (GOLDSHIMIDT; PASSOS; PASSOS, 2021).

Os valores Vp e Vn , são, respectivamente, verdadeiros positivos e verdadeiros negativos, onde o modelo acertou a classificação. E os valores Fp e Fn são, respectivamente, falsos positivos e falsos negativos, onde o modelo errou a classificação. Sendo assim, será utilizada a seguinte equação para medir a acurácia do modelo (BEZERRA, 2017):

$$Ac = \frac{(Vp + Vn)}{(Vp + Vn + Fp + Fn)} \quad (3.3)$$

Por fim, em caso de igualdade de desempenho usando a Equação 3.3 para duas técnicas, a que será considerada como a que melhor performou será a técnica que tiver o maior percentual de Vp (SHAABAN et al., 2012).

4 EXPERIMENTOS E DISCUSSÕES

4.1 Desenvolvimento dos modelos

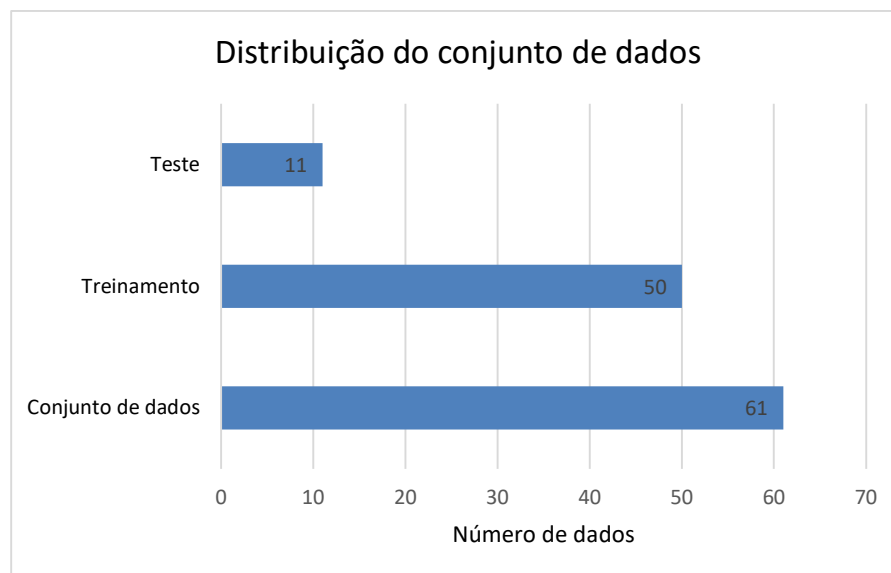
Para todos os experimentos realizados neste trabalho, foi utilizada a linguagem Python (ROSSUM, 1995) para construção dos modelos de aprendizado de máquina. O *software* utilizado para escrita do código foi o *Visual Studio*, ele é gratuito, portanto não representou nenhum impedimento. Além disso, foram utilizadas as bibliotecas Keras e Scikit-Learn (PEDREGOSA, 2011), para a etapa de pré-processamento de dados foram usadas as bibliotecas Numpy (OLIPHANT, 2016), Matplotlib (HUNTER, 2007) e Pandas (MCKINNEY, 2010).

4.2 Experimentos

4.2.1 Predição quantitativa

Todos os experimentos que irão compor essa parte do projeto tiveram como objetivo prever quantas assinaturas irão cancelar no próximo mês e, por isso, foram modelos de regressão. Todos também utilizaram o conjunto de dados estruturados no Quadro 1. Ao todo, ele possui 61 registros e sobre eles será aplicada a métrica de validação cruzada, com 80% dos dados para treinamento do modelo e 20% para teste, como mostra o Gráfico 2. A utilização do mesmo conjunto de dados e métricas de validação para todas as técnicas se justifica pela intenção de compará-los. Além disso, sobre os valores de teste, a métrica estatística de validação aplicada foi a MAPE.

Gráfico 2 - Distribuição do conjunto de dados para modelos de predição quantitativa.



Fonte: Próprio Autor.

4.2.1.1 *Random Forest*

Este experimento consiste na aplicação da técnica *Random Forest* e, a partir dos testes realizados, os valores que apresentaram melhor desempenho foram escolhidos para compor o experimento desta seção. Sendo eles a máxima profundidade de árvore e quantidade de árvores, como mostra o Quadro 3.

Quadro 3 – Parâmetros para a técnica *Random Forest*

Característica	Quantidade
Máxima profundidade	30
Quantidade de árvores	40

Fonte: Próprio Autor.

Para analisar o desempenho do modelo gerado, foram observados os valores da MAPE para o conjunto de dados de treinamento e para o conjunto de dados de teste, comparando-se os valores estimados e os valores reais. Um baixo valor para a média percentual do erro absoluto dos dados de treinamento seguindo de uma diferença ou semelhança entre esses valores para o conjunto

de treinamento e teste indicam se houve ou não *overfitting*, ou seja, se o modelo se especializou nos dados de treinamento, mas não apresenta um bom desempenho para dados nunca antes vistos, os quais simbolizam os dados de situações do mundo real.

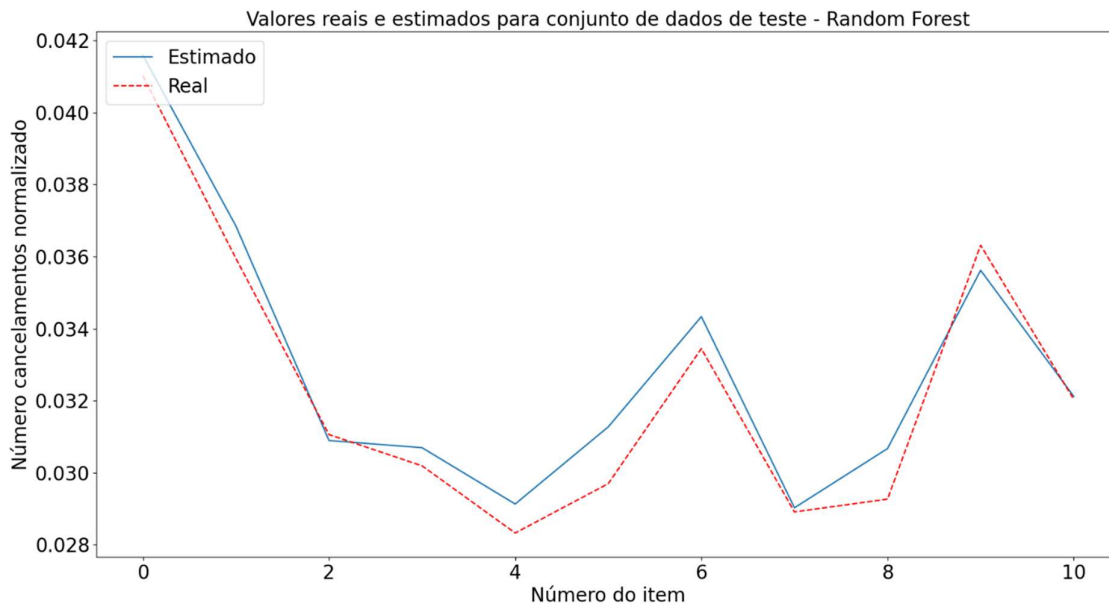
Sendo assim, o modelo apresentou um valor baixo para a MAPE sobre o conjunto de dados de treinamento e também para o conjunto de dados de teste, como mostra o Quadro 4. Por fim, o Gráfico 3 ilustra a semelhança entre as curvas geradas pelas saídas estimadas e reais do conjunto de dados de teste, onde o número do item representa um determinado mês.

Quadro 4 – Média percentual do erro absoluto – *Random Forest*

Conjunto de dados	MAPE
Treinamento	6,07
Teste	2,73

Fonte: Próprio Autor.

Gráfico 3 - Valores reais e estimados para conjunto de dados de teste - *Random Forest*.

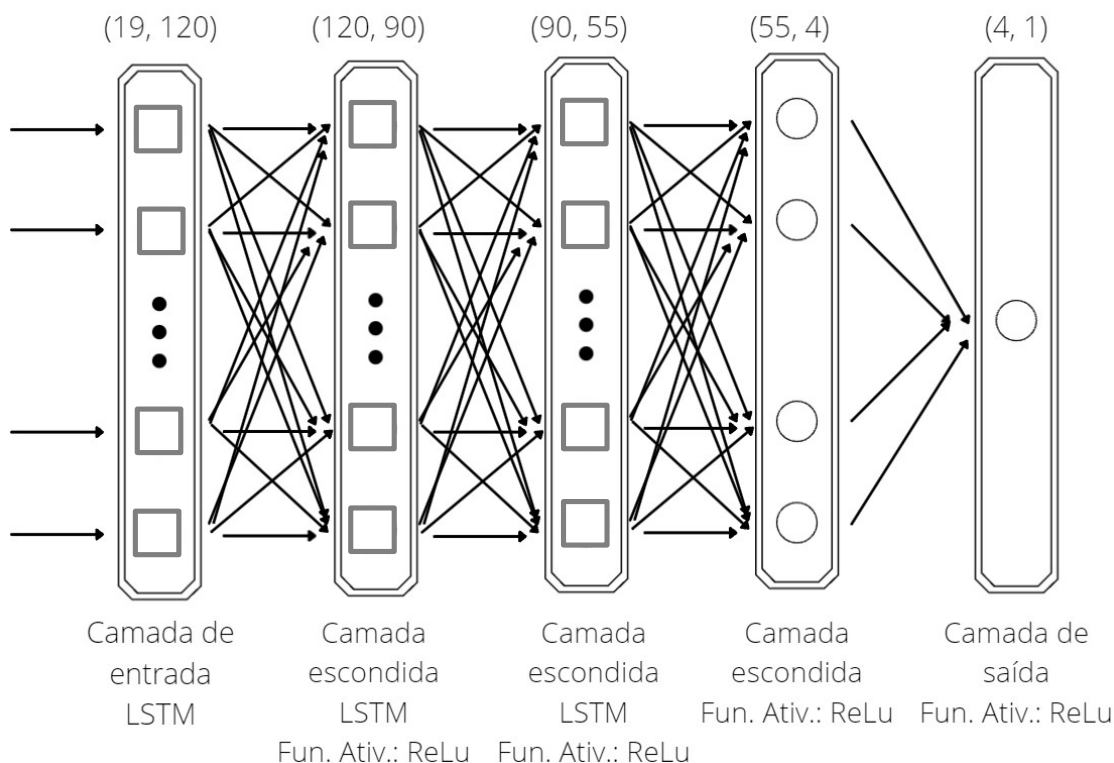


Fonte: Próprio Autor.

4.2.1.2 Feed Forward em conjunto com Long Short-Term Memory

O segundo experimento realizado foi com a aplicação da técnica de rede neural *Feed Forward* em conjunto com *Long Short-Term Memory*, também com vários testes realizados e observação do desempenho deles. A Figura 11 ilustra a configuração dela.

Figura 11 - Estrutura da rede neural *Feed Forward* com LSTM



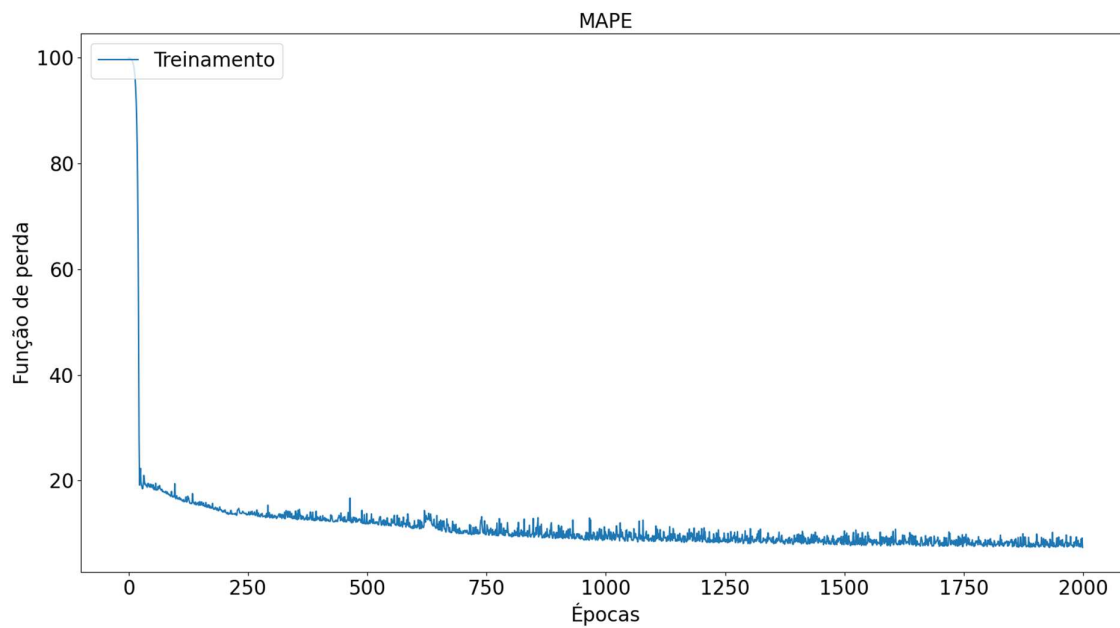
Fonte: Próprio Autor.

Com o processo de treinamento e de teste aplicados, a rede neural performou com um valor de 7,77% de MAPE para o conjunto de treinamento, como mostra o Quadro 5, ou seja, quando se comparou o resultado estimado com o real. Além disso, apresentou um valor mais baixo para a MAPE do conjunto de dados de teste, 5,19%, como mostra o Gráfico 4. Esse acontecimento comprova que não ocorreu *overfitting*, em outras palavras, o modelo não se especializou somente no conjunto de dados de treinamento, ele apresentou a mesma performance para os dois conjuntos, assim indicando que situações nunca antes vistas serão bem avaliadas.

Quadro 5 – Média percentual do erro absoluto – *Feed Forward* com LSTM

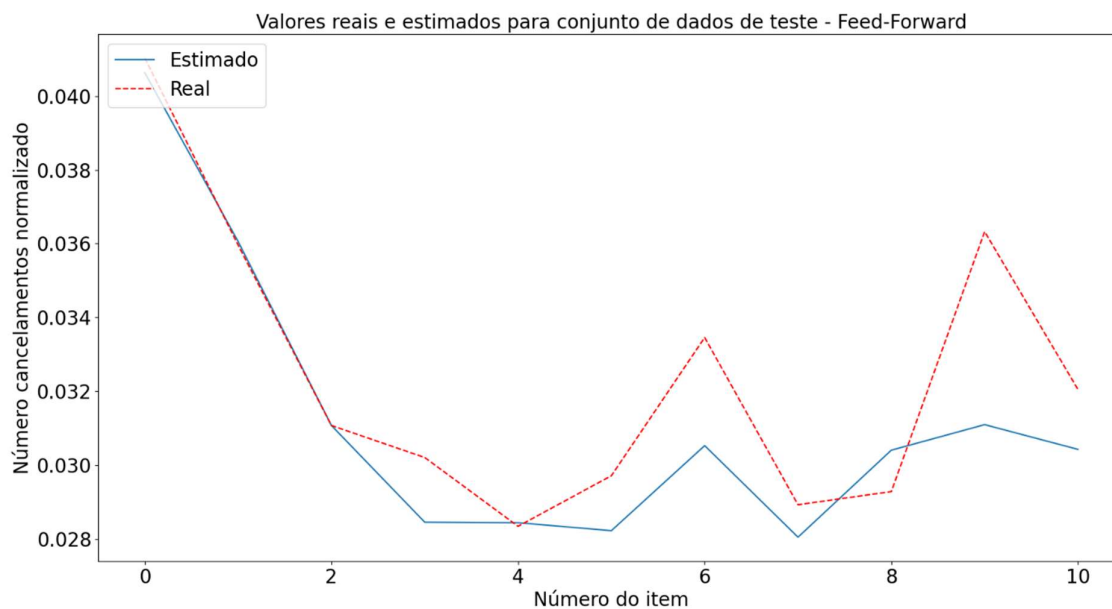
Conjunto de dados	MAPE
Treinamento	7,77
Teste	5,19

Fonte: Próprio Autor.

Gráfico 4 - MAPE - *Feed Forward* com LSTM

Fonte: Próprio Autor.

Em seguida, foi analisado o desempenho dos dados de teste e como a curva gerada pelo modelo para esses valores se comporta, o Gráfico 5 ilustra esse aspecto.

Gráfico 5 - Valores reais e estimados para o conjunto de teste - *Feed Forward* com LSTM

Fonte: Próprio Autor.

4.2.1.3 Comparação dos resultados

Com todos os experimentos para predição da quantidade de cancelamentos realizados, analise-se o desempenho das técnicas sobre o conjunto de dados de teste por meio dos valores da métrica escolhida para esta etapa, a MAPE, como mostra o Quadro 6.

Quadro 6 – Comparação do desempenho das técnicas

Técnica aplicada	MAPE
<i>Random Forest</i>	2,73
<i>Feed Forward</i> com LSTM	5,19

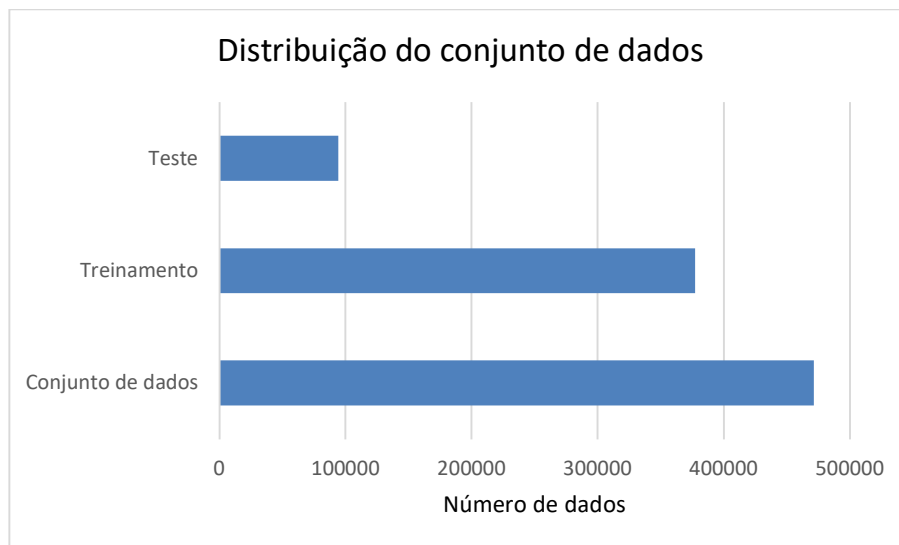
Fonte: Próprio Autor.

4.2.2 Predição qualitativa

Todos os experimentos que irão compor essa parte do projeto terão como objetivo prever quais assinaturas irão cancelar no próximo mês e por isso serão gerados modelos de classificação. Todos também irão utilizar o conjunto de dados estruturados no Quadro 2. Ao todo, ele possui 471.408 registros e sobre eles será aplicada a métrica de validação cruzada,

com 80% dos dados para treinamento do modelo e 20% para teste, como mostra o Gráfico 6. A utilização do mesmo conjunto de dados e métricas de validação para todas as técnicas se justifica pela intenção de compará-los. Além disso, sobre os valores de teste, a métrica estatística aplicada foi explicada a explicitada na Equação 3.3 para medir o desempenho do modelo.

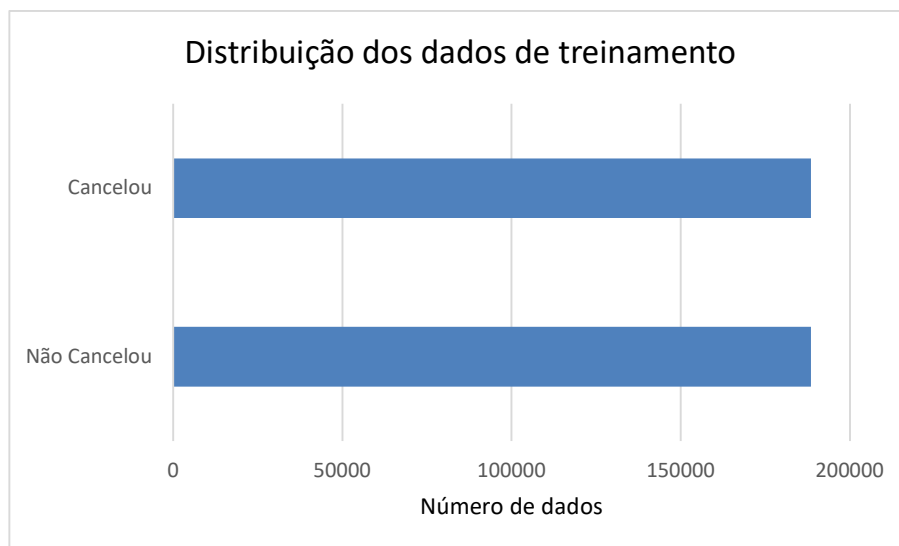
Gráfico 6 - Distribuição dos dados predição qualitativa.



Fonte: Próprio Autor.

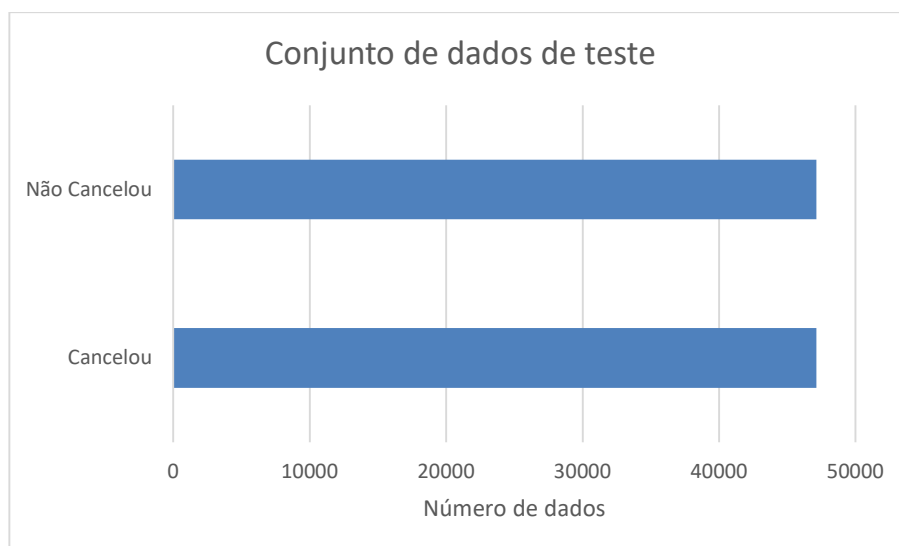
Por fim, para o conjunto de dados de treinamento, como mostra o Gráfico 7, e de teste, como mostra o Gráfico 8, será utilizada a mesma quantidade de dados referentes à assinaturas que cancelaram e que não cancelaram, ou seja, os conjuntos de dados serão balanceados. Isso será realizado para que o modelo não tenha nenhuma tendência à sua predição.

Gráfico 7 - Conjunto de dados de treinamento.



Fonte: Próprio Autor.

Gráfico 8 - Conjunto de dados de teste.



Fonte: Próprio Autor.

4.2.2.1 *Random Forest*

Neste experimento, é aplicada a técnica *Random Forest* para predizer quais assinaturas irão cancelar no próximo mês. Os valores de configuração da técnica foram escolhidos de acordo com os testes realizados e observação do seu desempenho, entre eles estão a quantidade de árvores e máxima profundidade, como mostra o Quadro 7.

Quadro 7 – Parâmetros pra a técnica *Random Forest*

Característica	Quantidade
Máxima profundidade	50
Quantidade de árvores	100

Fonte: Próprio Autor.

Para medir o desempenho do modelo gerado, foi analisado o valor encontrado pela Equação 3.3 com relação aos valores reais e estimados nos conjuntos de dados de treinamento e teste. Como os valores encontrados na Equação 3.3 para os dados de treinamento e de teste foram significativamente diferentes, como mostra o Quadro 8, conclui-se que houve *overfitting*.

Quadro 8 - Acurácia para conjunto de treinamento e teste - *Random Forest*

Técnica	Acurácia percentual
Treinamento	98,32
Teste	80,57

Fonte: Próprio Autor.

Outro fator observado é o percentual referente ao V_p , V_n , F_p e F_n sobre o valor total de dados de teste, como mostra o Quadro 9. O valor percentual de V_p se refere à taxa de ocorrência dos verdadeiros positivos, V_n à taxa de ocorrência dos verdadeiros negativos, F_p à quantidade de falsos positivos sobre o número total de dados de teste e F_n à taxa de ocorrência dos falsos negativos. Sendo portanto a taxa de V_p a mais importante, pois diz respeito às predições corretas sobre as assinaturas que irão cancelar no próximo mês.

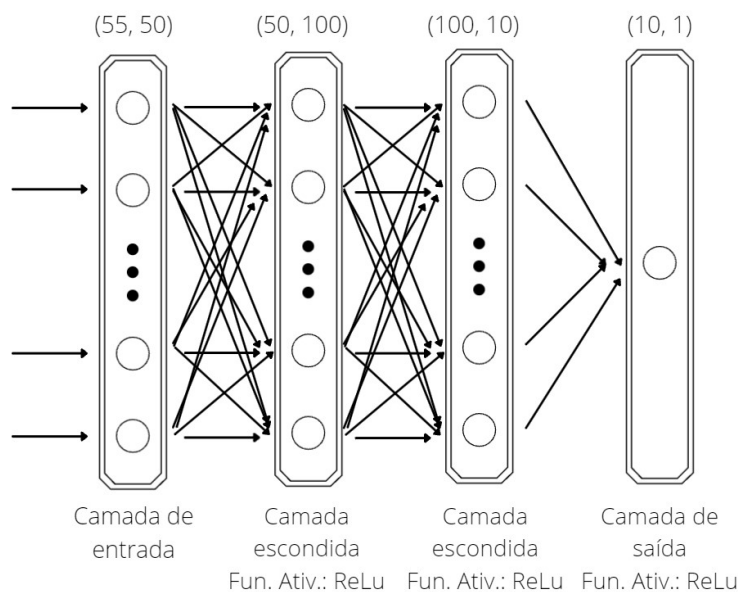
Quadro 9 - Matriz confusão - *Random Forest*.

	Valores Reais	Valores Preditos	
		Não Cancelou	Cancelou
Random Forest	Não Cancelou	38.045 / 47.141 80,70%	9.096 / 47.141 19,30%
	Cancelou	9.217 / 47.141 19,55%	37.924 / 47.141 80,45%

Fonte: Próprio Autor.

4.2.2.2 *Feed Forward*

Este experimento consiste na aplicação da técnica *Feed Forward* para prever quais assinaturas serão canceladas no mês seguinte. Após alguns modelos gerados, foi escolhida a melhor configuração da rede neural de acordo com o resultado do seu desempenho. A configuração diz respeito à quantidade de neurônios e função de ativação, como ilustra a Figura 12.

Figura 12 - Estrutura rede neural *Feed Forward*.

Fonte: Próprio Autor.

Para medir o desempenho do modelo, foi observado o valor encontrado pela métrica de validação sobre os valores estimados e reais dos conjuntos de dados de treinamento e teste, como mostra o Quadro 10, bem como os valores das taxas sobre V_p , V_n , F_p , e F_n pela matriz de confusão, como mostra o Quadro 11.

Quadro 10 - Acurácia para conjunto de treinamento e teste - Feed Forward

Técnica	Acurácia percentual
Treinamento	75,89
Teste	76,32

Fonte: Próprio Autor.

Quadro 11 - Matriz de confusão - Feed Forward

	Valores Reais	Valores Preditos	
		Não Cancelou	Cancelou
<i>Feed Forward</i>	Não Cancelou	31.543 / 47.141 66,91%	15.816 / 47.141 33,55%
	Cancelou	6.506 / 47.141 14,26%	40.416 / 47.141 85,73%

Fonte: Próprio Autor.

4.2.2.3 Comparação dos resultados

Com todos os experimentos para predição de quais assinaturas irão cancelar no mês seguinte, analisa-se o desempenho das técnicas sobre os conjuntos de dados de teste por meio dos valores das métricas escolhida para este projeto, a acurácia percentual (Equação 3.3) e o percentual de V_p , como mostra o Quadro 12.

Quadro 12 - Comparação de desempenho das técnicas

Técnica aplicada	Acurácia	Percentual de Vp
<i>Random Forest</i>	80,57	80,45
<i>Feed Forward</i>	76,32	85,73

Fonte: Próprio Autor.

5 CONCLUSÕES

Este projeto teve como objetivo prever a quantidade de cancelamentos mensais de assinaturas mensais de vinho da empresa Wine, previsão quantitativa, e também prever quais assinaturas serão canceladas no mês posterior, previsão qualitativa, para isso, foram aplicadas 2 técnicas de aprendizado de máquina de regressão e 2 de classificação respectivamente. Além disso, este projeto buscou apresentar todo o processo para geração desses modelos, desde à extração dos dados até alimentação dos modelos com os conjuntos de dados extraídos e pré-processados.

Para a previsão quantitativa, foram observadas dificuldades para ambos os modelos estabilizarem seus resultados, principalmente pela quantidade de dados. Modelos de aprendizado de máquina normalmente necessitam de muitos dados para se tornarem precisos e identificarem padrões, portanto isso tornou os resultados extremamente sensíveis aos conjuntos escolhidos para treinamento e teste.

Apesar disso, as técnicas apresentaram bons resultados, como demonstrado no Capítulo 4, a técnica que apresentou melhor performance foi a *Random Forest*, resultado determinado através da função de perda escolhida para esta etapa, a MAPE. Os valores foram 2,73% para a *Random Forest* e 5,19% para a *Feed Forward* com LSTM nessa métrica. Não foi observado o *overfitting* nessa comparação, pois os dois modelos apresentaram resultados bons nesse aspecto.

Para a previsão qualitativa, a técnica *Random Forest* apresentou um bom resultado enquanto que a rede neural apresentou um resultado razoável sobre o valor de acurácia, 80,57% e 76,32% respectivamente. Entretanto, o modelo de *Random Forest* gerado apresentou *overfitting*, quando o modelo performa muito bem para os dados de treinamento e não tão bem para os dados de teste, enquanto que a rede neural não, além disso, tivemos um melhor resultado na rede neural sobre os V_p , ou seja, sobre a previsão das assinaturas que cancelaram no mês seguinte, 85,73%, enquanto que a outra apresentou 80,45%. Por este ser o principal objetivo deste projeto, é também o principal critério para escolha do melhor resultado.

Para trabalhos futuros, sobre a previsão quantitativa, é importante aumentar a quantidade de informações analisadas nos meses, uma vez que essa é a única alternativa para fornecer mais informações para a geração dos modelos, como sugestão fica a quantidade de indicações de

clientes para clientes que ocorreram no mês anterior e o valor gasto médio pelos assinantes. Outro ponto importante é a aplicação de outras técnicas não lineares para comparação de desempenho.

Sobre a previsão qualitativa, também é interessante trazer mais informações, principalmente pela correlação existente entre diferentes comportamentos dos clientes, como sugestão, ficam todas as informações que refletem um comportamento próximo à ação de cancelar uma assinatura, bem como aplicar mais técnicas para efeito de comparação. Por fim, é muito importante que se aplique um método de *clustering* sobre os Vp para que se detecte os padrões entre eles e direcione assim ações para retenção.

REFERÊNCIAS BIBLIOGRÁFICAS

AHMAD, A. K; JAFAR, A; ALJOUAAA, K. **Customer churn prediction in telecom using machine learning in big data platform.** Journal of Big Data, v. 6, p. 28, 2019.

ALSABTI, K. RANKA, S. SINGH, V. **Clouds:** a decision tree classifier for large databases. University of Florida, 1998. 34p. Disponível em: <https://www.aaai.org/Papers/KDD/1998/KDD98-001.pdf>. Acesso em 11 set. 2021.

AUGUSTIN, N. H. BUCKLAND, S. T. BURNHAM, K. P. **Model selection:** An integral part of inference. Biometrics, v. 53, p. 603-618, 1997.

BAO, W.; YUE, J.; RAO, Y. **A deep learning framework for financial time series using stacked autoencoders and long-short term memory.** PLoS ONE, v. 12, p. 1–24, jul. 2017.

BEZERRA, A. A. **Predição de desempenho de estudantes em ambientes virtuais de aprendizagem.** 2017. Dissertação (Mestrado) – Universidade Federal do Ceará, Campinas, 2017.

BRAGA, A. P. CASTRO, C. L. **Aprendizado de máquina com conjunto de dados desbalanceados.** SBA: Controle e automação v. 11, p. 169-198, 2011. Disponível em: <https://www.scielo.br/j/ca/a/pXMZjzHJcJtkLVYLLDHHTxw/?lang=pt>. Acesso em 12 set. 2021.

BREIMAN, L. FRIEDMAN, J. H. OLSHEN, R. A. STONE, C. J. **Classification and regression trees.** 1. ed. New York: CRC Press, 1984.

BREIMAN, L. **Bagging predictors.** Machine Learning, v. 24, n. 2, p.123-140, 1996.

BREIMAN, L. **Random forests.** Machine Learning, v. 45, n.1, p.5-32, 2001.

BÜHLMANN, P. YU, B. **Analysing bagging.** The annals of statistics. v. 30, 2002.

CARVALHO, A. C. P. L. F. FACELI, K. LORENA, A. C. GAMA, J. **Inteligência Artificial: Uma abordagem de Aprendizado de Máquina.** 1. ed. LTC, 2011.

DANTAS, D. DONADIA, E. D. **Comparação entre as técnicas de regressão logística, árvore de decisão, bagging e random forest aplicadas à um estudo de concessão de crédito.** 2013. Trabalho de Conclusão de Curso (Graduação em Estatística) – Departamento de Estatística, Universidade Federal do Paraná, Curitiba, 2013.

DUDA, R. O. HART, P. E. STORK, D. G. **Pattern classification**. 2. ed. New York, 2001.

DUDA, R. O. HART, P. E. STORK, D. G. **Pattern classification**. 2. ed. John Wiley & Sons, 2012.

EFRON, B. TIBSHIRANI, R. J. **An introduction to the Bootstrap**. 1. ed. CRC Press, New York, 1993.

ELLIOTT, D. L. **A better activation function for artificial neural networks**. Technical Research Report. Maryland, 1993. Disponível em: https://www.researchgate.net/publication/277299531_A_better_Activation_Function_for_Artificial_Neural_Networks. Acesso em 11 set. 2021.

FERNANDEZ, C. M. **Uma arquitetura de *software* para descoberta de regras de associação multidimensional, multinível e de outliers em cubos OLAP: um estudo de caso com os algoritmos Apriori e FP-Growth**. 2010. Dissertação (Mestrado) – Centro de Informática, Universidade Federal de Pernambuco, Recife, 2010.

FLAUSINO, R. A. DA SILVA, I. N. SPATTI, D. H. **Redes neurais artificiais para engenharia e ciências aplicadas: Fundamentos teóricos e aspectos práticos**. 2. ed. Artliber, São Paulo, 2010.

GARCÍA, S. HERRERA, F. LUENGO, J. **Data preprocessing in data mining**. Springer, 2014.

GOLDSHIMIDT, R; PASSOS, E; PASSOS, E. **Data Mining: Conceitos, Técnicas, Algoritmos, Orientações e Aplicações**. 5. ed. [S.l: s.n.], 2021.

HALMENSCHLAGER, C. **Um algoritmo para indução de árvores e regras de decisão**. 2002. Dissertação (Pós-Graduação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

HAN, J. KAMBER, M. **Data mining: concepts and techniques**. 2. ed. Morgan Kaufmann, 2000.

HAYKIN, S. **Neural networks and learning machines**. 3. ed. Pearson Prentice Hall, New Jersey, 2008.

HINTON, G. E. RUMELHART, D. WILLIAMS, R. J. **Learning representations by back-propagating errors**. Neurocomputing: foundations of research. MIT Press. p. 696-699 Cambridge, 1998.

HOCHREITER, S. SCHMIDHUBER, J. **Long Short-Term Memory**. *Neural Computation*. v. 9, 1997.

HUNTER, J. D. **Matplotlib**: A 2d graphics environment. 3. ed. , 2007. v. 9.

MCKINNEY, Wes. *Data structures for statistical computing in Python*. jan. 2010, Austin: [s.n.], jan. 2010. p. 51–56.

IZQUIERDO, I. **Memoria**. 2. ed. Porto Alegre: Artmed, 2011.

KELLER, P. E. PRIDDY, P. E. **Artificial neural networks**: An introduction. SPIE, 2005.

KNORR, E. **Outliers and data mining**: finding exceptions in data. 2002. Tese (Doutorado em Filosofia) – Department of Computer Science, University of British Columbia, 2002.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, Nature Publishing Group. v. 521, n. 7553, p. 436–444, 2015.

LIMA, J. G. B. **Uma aplicação de impacto social com aprendizado de máquina**. 2014. Trabalho de Conclusão de Curso (Graduação em Sistemas de Informação) – Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Recife, 2014.

MASCARENHAS, T. A. T. **Aplicação de algoritmos de aprendizado de máquina na classificação de conhecimentos especializados de professores de Física**. 2020. Dissertação (Mestrado) – Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso, Cuiabá, 2020.

MESQUITA, P. S. B; NOGUEIRA, R. T. **Um modelo estatístico para gestão de programas de pós-graduação**. *Gestão da Produção, Operações e Sistemas*, v. 2, p. 173–186, 2015.

MONTANARI, R. **Detecção e classificação de objetos em imagens para rastreamento de veículos**. 2015. Dissertação (Mestrado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2015.

OLAH, C. **The Unreasonable Effectiveness of Recurrent Neural Networks**. 2015. Disponível em: <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>. Acesso em: 20 jul. 2021.

OLIPHANT, T. **A guide to NumPy**. 2. ed. [S.l: s.n.], 2016. v. 1.

OLIVEIRA, A. R. **Comparação de algoritmos de aprendizado de máquina para construção de modelos preditivos de diabetes não diagnosticado**. 2016. Dissertação (Mestrado) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.

OPTIZ, D. MACLIN, R. **Popular Ensemble Methods: An Empirical Study**. *Journal Of Artificial Intelligence Research*, v. 11, p. 169-198, 1999.

PEDREGOSA, F. **Scikit-Learn: Machine Learning in Python**. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

PEREIRA, C. R. **Aprendizado de máquina aplicado ao auxílio do diagnóstico da doença de Parkinson**. 2017. Tese (Doutorado em Ciência de Computação) – Universidade Federal de São Carlos, São Carlos, 2017.

PEREIRA, D. M. **Uso do padrão OIM de metadados no suporte às transformações em ambientes Data Warehouse**. 2000. Dissertação (Mestrado) – Instituto de Matemática e Núcleo de Computação Eletrônica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2000.

PONTES, F. J. **Projeto otimizado de redes neurais artificiais para predição de rugosidade em processos de usinagem com a utilização de metodologia de projetos de experimentos**. 2011. Tese (Doutorado em Engenharia Mecânica) – Faculdade de Engenharia do Campus de Guaratinguetá, Universidade Estadual Paulista, Guaratinguetá, 2011.

POLAT, C. **The Role of Forecasting and Its Potential for Functional Management: A Review from the Value-Chain Perspective**. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, v. 9, 2007.

PRATI, R. C. **Novas abordagens em aprendizado de máquina para geração de regras, classes desbalanceadas e ordenação de casos**. 2006. Tese (Doutorado em Ciência de Computação e Matemática Computacional) – Universidade de São Paulo, São Carlos, 2006.

RODRIGUES, B. G. D. **Uso de árvore de decisão para avaliação da segurança estática em tempo real de sistemas elétricos de potência.** 2014. Dissertação (Pós-Graduação) – Instituto de Informática, Universidade Federal do Pará, Belém, 2014.

ROSSUM, Guido van. **Python Reference Manual.** Amsterdam: Centre for Mathematics and Computer Science, 1995.

RUSSEL, S. NORVIG, P. **Inteligência Artificial.** 3. ed. Rio de Janeiro, 2013.

SAMUEL, A. L. *Some Studies in Machine Learning Using the Game of Checkers.* *IBM Journal of Research and Development*, v. 3, n. 3, p. 210-229, 1959.

SANTOS, A. P. **Análise e comparação do desempenho de diferentes arquiteturas de redes neurais artificiais profundas aplicadas à traigem de lesões de pele.** 2019. Dissertação (Mestrado) - Universidade Presbiteriana Mackenzie, São Paulo, 2019.

SANTOS, G. S. M. **Aplicação de redes neurais para diagnóstico de poluição em isoladores de alta tensão.** 2008. Dissertação (Mestrado) – Centro de Tecnologia e Geociências, Universidade Federal de Pernambuco, Recife, 2008.

SCALABRIN, M. H. **Aplicação de redes neurais artificiais profundas na detecção de placas de pare.** 2019. Trabalho de Conclusão de Curso (Especialização em Métodos Matemáticos Aplicados) – Departamento Acadêmico de Informática, Universidade Tecnológica Federal do Paraná, Francisco Beltrão, 2019.

SHAABAN, E. HELMY, Y. K. KHEDR, A. E. NASR, M. M. **A proposed churn prediction model.** *International Journal of Engineering*, v. 2, n. 4, p. 693–697, jun. 2012.

SILVA, C. L. **Aprendizado de máquina com treinamento continuado aplicado à previsão de demanda de curto prazo: o caso do restaurante universitário da Universidade Federal de Uberlândia.** 2019. Dissertação (Mestrado) – Universidade Federal de Uberlândia, Uberlândia, 2019.

SILVA, N. C. **Análise e síntese de estratégias de aprendizado para redes neurais artificiais.** 1998. Dissertação (Mestrado) – Departamento de Engenharia de Computação e Automação Industrial, Universidade Estadual de Campinas, Campinas, 1998.

SOUZA, C. O. F. **Desenvolvimento de aplicações ETL como uma proposta para redução de custos em projeto de data warehouse.** 2003. Dissertação (Mestrado) – Departamento de Eletrônica e Sistemas, Universidade Federal de Pernambuco, Recife, 2003.

TANURE, C. M. **Uma arquitetura de *software* para descoberta de regras de associação multidimensional, multinível e de outliers em cubos OLAP: um estudo de caso com os algoritmos Apriori e FP-Growth.** 2010. Dissertação (Mestrado) – Centro de Informática, Universidade Federal de Pernambuco, Recife, 2010.

THEODORIDIS, S. KOUTROUMBAS, K. **Clustering: Basic concepts.** Pattern recognition. p. 483-516, 2006.

TURING, A. M. ***Computing machinery and intelligence-am turing.*** Mind, Oxford University Press *on behalf of the* Mind Association. v. 59, n. 236, p. 433, 1950.

VASCONCELOS, J. H. **Investigações sobre métodos de classificação para uso de termografia de mama.** 2017. Dissertação (Mestrado) – Centro de Tecnologia e Geociências, Universidade Federal de Pernambuco, Recife, 2017.