



UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA E CONTABILIDADE
DEPARTAMENTO DE ADMINISTRAÇÃO
CURSO DE CIÊNCIAS ATUARIAIS

JOÃO PEDRO DE CASTRO LOPES DA SILVEIRA

O IMPACTO FINANCEIRO DA PANDEMIA DE COVID-19 NO MERCADO DE
ODONTOLOGIA BRASILEIRO: UM ESTUDO DE CASO NO GRUPO
ODONTOPREV

FORTALEZA-CE

2023

JOÃO PEDRO DE CASTRO LOPES DA SILVEIRA

O IMPACTO FINANCEIRO DA PANDEMIA DE COVID-19 NO MERCADO DE
ODONTOLOGIA BRASILEIRO: UM ESTUDO DE CASO NO GRUPO ODONTOPREV

Monografia apresentada ao Curso de Ciências Atuariais do Departamento de Administração da Universidade Federal do Ceará, como requisito parcial para obtenção do Título de Bacharel em Ciências Atuariais.

Orientadora: Prof^ª. Ms. Alana Katielli Nogueira Azevedo.

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S588i Silveira, João Pedro de Castro Lopes da.
O IMPACTO FINANCEIRO DA PANDEMIA DE COVID-19 NO MERCADO DE ODONTOLOGIA
BRASILEIRO : UM ESTUDO DE CASO NO GRUPO ODONTOPREV / João Pedro de Castro Lopes da
Silveira. – 2023.
42 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Faculdade de Economia,
Administração, Atuária e Contabilidade, Curso de Ciências Atuariais, Fortaleza, 2023.
Orientação: Profa. Ma. Alana Katielli Nogueira Azevedo.
1. Pandemia. 2. Odontologia. 3. Power BI. I. Título.

CDD 368.01

JOÃO PEDRO DE CASTRO LOPES DA SILVEIRA

O IMPACTO FINANCEIRO DA PANDEMIA DE COVID-19 NO MERCADO DE
ODONTOLOGIA BRASILEIRO: UM ESTUDO DE CASO NO GRUPO ODONTOPREV

Monografia apresentada ao Curso de Ciências
Atuariais do Departamento de Administração da
Universidade Federal do Ceará, como requisito
parcial para obtenção do Título de Bacharel em
Ciências Atuariais.

Aprovada em: 08/12/2023.

BANCA EXAMINADORA

Prof^ª. Ms. Alana Katielli Nogueira Azevedo (Orientadora)
Universidade Federal do Ceará (UFC)

Prof^ª. Dra. Alane Siqueira Rocha
Universidade Federal do Ceará (UFC)

Prof. Dr. Daniel Tomaz de Sousa
Universidade Federal do Ceará (UFC)

Aos meus avôs e avós, que sempre olharam e oraram por mim e que, onde quer que estejam, sempre iluminam e abençoam os meus dias.

RESUMO

O presente trabalho aborda como tema o impacto do cenário imposto pela pandemia do vírus da Covid-19 sobre o mercado de odontologia brasileiro, avaliando graficamente e estatisticamente as variáveis envolvidas no problema. O principal objetivo foi dar luz ao mercado de odontologia brasileiro dentro do cenário da ciência atuarial, assim como avaliar as variáveis econômicas e financeiras do grupo, e avaliar se as variáveis relativas à pandemia oferecem algum tipo de interferência nos números referentes ao grupo no recorte temporal pandêmico. Para alcançar o objetivo do trabalho, usou-se duas ferramentas de análise de dados: o *Power BI*, que se utiliza de *Business Intelligence* no seu escopo; e o *Orange Data Mining*, que é um *software* de mineração e análise de dados, utilizado principalmente em aprendizado de máquina (*Machine Learning*). Na primeira ferramenta, foi realizada uma análise gráfica para entender o comportamento das variáveis do grupo Odontoprev; na segunda ferramenta, foi feita uma análise estatística, comparando modelos, avaliando qual melhor se ajusta aos dados, e para assim, através do coeficiente de determinação, entender o impacto que as variáveis referentes à pandemia exercem sobre as variáveis do grupo Odontoprev. Feita a análise, concluiu-se que houve sim impacto do período pandêmico no crescimento dos números referentes ao crescimento econômico do grupo Odontoprev no período em questão, através de análises gráficas e análise de coeficiente de determinação.

Palavras-chave: pandemia; odontologia; *Power BI*.

ABSTRACT

This work addresses the impact of the scenario imposed by the Covid-19 virus pandemic on the Brazilian dentistry market, graphically and statistically evaluating the variables involved in the problem. The main objective was to shed light on the Brazilian dentistry market within the actuarial science scenario, as well as evaluate the economic and financial variables of the Odontoprev group, and evaluate whether variables related to the pandemic offer any type of interference in the numbers referring to the group in the pandemic time frame. To achieve the objective of the work, two data analysis tools were used: Power BI, which uses Business Intelligence in its scope; and Orange Data Mining, which is data mining and analysis software, mainly used in machine learning. In the first tool, a graphical analysis was carried out to understand the behavior of the variables in the Odontoprev group; in the second tool, a statistical analysis was carried out, comparing models, evaluating which best fits the data, and thus, through the coefficient of determination, understanding the impact that the variables related to the pandemic have on the variables of the Odontoprev group. After the analysis, it was concluded that there was an impact of the pandemic period on the growth of numbers referring to the economic growth of the Odontoprev group in the period in question, through graphical analyzes and coefficient of determination analysis.

Keywords: pandemic; dentistry; Power BI.

LISTA DE FIGURAS

Figura 1	– Etapas que compõem o Business Intelligence de forma detalhada.....	15
Figura 2	– Exemplo de Matriz de Confusão.....	19
Figura 3	– Resumo das escolhas das variáveis para compor os nós de uma árvore aleatória.....	21
Figura 4	– Imagem que exemplifica uma Rede Neural <i>Feedforward</i> (FNN).....	24
Figura 5	– Funcionamento da técnica de Validação cruzada.....	27
Figura 6	– Representação da tela de análises do <i>Orange Data Mining</i>	32

LISTA DE GRÁFICOS

Gráfico 1 – Área sob a curva ROC.....	
Gráfico 2 – Reta de regressão linear aplicada sobre dados.....	22
Gráfico 3 – Classificação de objeto antes/depois do treinamento dos dados pelo algoritmo KNN.....	25
Gráfico 4 – Evolução da quantidade de vidas do grupo Odontoprev, por ano.....	29
Gráfico 5 – Evolução da receita líquida do grupo Odontoprev, por ano.....	29
Gráfico 6 – Evolução do lucro líquido do grupo Odontoprev, por ano.....	30
Gráfico 7 – Evolução da sinistralidade do grupo Odontoprev, por ano.....	31
Gráfico 8 – Evolução do ticket médio do grupo Odontoprev, por ano.....	31
Gráfico 9 – Casos e óbito por covid-19, por ano.....	33

LISTA DE TABELAS E QUADROS

Quadro 1 – Tipos de neurônios, suas características, vantagens e desvantagens.....	23
Tabela 1 – Tabela de valores de R^2 para cada modelo.....	35

SUMÁRIO

1	INTRODUÇÃO	11
2	REFERENCIAL TEÓRICO.....	12
2.1	Odontoprev: um breve histórico.....	12
2.2	Business Intelligence.....	13
2.3	Power BI.....	15
2.4	Orange Data Mining.....	16
2.5	Problemas de regressão e classificação.....	17
2.5.1	<i>Regressão</i>.....	17
2.5.2	<i>Classificação</i>.....	18
2.6	Modelos e métodos avaliados.....	20
2.6.1	<i>Random Forest</i>.....	20
2.6.2	<i>Linear Regression (Regressão Linear)</i>.....	21
2.6.3	<i>Neural Network (Rede Neural)</i>.....	22
2.6.4	<i>KNN (K-vizinhos mais próximos)</i>.....	24
2.6.5	<i>Cross-Validation (Validação Cruzada)</i>.....	26
3	APLICAÇÃO METODOLÓGICA.....	28
3.1	Variáveis utilizadas.....	28
3.2	Análise no <i>Power BI</i>.....	28
3.3	Análise no <i>Orange Data Mining</i>.....	32
4	ANÁLISE DOS RESULTADOS.....	34
4.1	Resultados da análise no <i>Power BI</i>.....	34
4.2	Resultados da análise no <i>Orange Data Mining</i>.....	34
5	CONSIDERAÇÕES FINAIS.....	36
	REFERÊNCIAS.....	37

1 INTRODUÇÃO

O presente trabalho busca analisar o impacto financeiro no mercado de odontologia brasileiro no período em que o país (e o mundo) esteve acometido pela crise sanitária imposta pela Covid-19. Para isso, tomou-se como estudo de caso o grupo Odontoprev.

O intuito principal com a análise realizada é entender e dar luz ao mercado odontológico brasileiro, visto que, dentro do contexto da ciência atuarial, não há um volume tão considerável de trabalhos ou estudos que abordem tal mercado em específico. De acordo com Aveni (2020), o mercado de saúde como um todo sofreu e sofre diversos impactos referentes ao cenário imposto pela pandemia, que são reflexos da má gestão no atendimento dos pacientes, as falhas do mercado de saúde, a falta de planejamento para lidar com crises, ausência de previsões e análises de risco. Na busca por resultados, se deseja entender qual o impacto sofrido pelo mercado dentro do período relativo ao cenário de crise sanitária enfrentado nos anos de pandemia.

A metodologia utilizada na análise consiste em avaliar variáveis de mercado como quantidade de vidas da companhia no período, receita líquida, lucro líquido, sinistralidade e ticket médio, assim como variáveis relacionadas a mortalidade e letalidade do vírus da Covid-19, relacioná-las e entender o real impacto que o período citado gerou em cada uma delas.

O estudo em questão visa trazer a luz um contexto do mercado odontológico que, como dito anteriormente, não é tão abordado pela academia. Desta forma, fazendo uma análise gráfica, através do *Power BI*, e analisando estatisticamente os comportamentos das variáveis em questão, utilizando o *Orange Data Mining*, pode-se entender quais os impactos de tal mercado e avaliar se a pandemia realmente gerou grandes variações econômicas, trazendo o caso prático como espelho do mercado como um todo, visto o tamanho e relevância da companhia estudada.

Considerando esta introdução, o texto é estruturado em cinco capítulos. O Capítulo 2 contextualiza a empresa caso de estudo, como também expõe as definições necessárias dos métodos a serem utilizados. O Capítulo 3 se concentra na metodologia, detalhando os procedimentos metodológicos empregados na pesquisa, fornecendo uma visão detalhada da abordagem utilizada. O capítulo 4 apresenta a análise dos resultados da pesquisa. Finalmente, o Capítulo 5 oferece considerações finais.

2 REFERENCIAL TEÓRICO

2.1 Odontoprev: um breve histórico

Em 1987, surgiu, a partir de um desejo de cirurgiões-dentistas, a Odontoprev. Os fundadores da empresa queriam oferecer um serviço odontológico de qualidade, levando em conta a crescente demanda de clientes.

Passaram-se dez anos, e em 1997, após um grande crescimento orgânico da empresa, ela consolidou-se como líder de mercado. A partir de então, a Odontoprev iniciou um período de aquisição de novas empresas, que se deu pela compra da Unidont, aumentando sua capacidade de atendimento e fortalecendo a marca. Em 1998, criou-se o conselho administrativo da companhia. Logo após a aquisição da empresa Unidont, deu-se início o apoio ao instituto Ronald McDonald, que oferece suporte às crianças que lutam contra o câncer.

No ano 2000, houve um marco na trajetória da Odontoprev: deu-se início a parceria com a Unimed Belo Horizonte. Com o seu crescimento, vindo através da aquisição de empresas concorrentes e de forma orgânica, a empresa atingiu a marca de um milhão de beneficiários.

Um dos grandes fatores de crescimento da companhia foi a abertura de capital na B3, tornando-se uma empresa de capital aberto no ano de 2006. Duas grandes aquisições vieram no ano seguinte: *DentalCorp* e Rede Dental, aumentando ainda mais a capilaridade da empresa. Mais três grandes compras ocorreram logo após esse período, em 2008: *Care Plus Dental*, Sepao e *Easy Dental Software*. Ainda no mesmo ano, a Odontoprev aderiu ao Pacto Global das Nações unidas.

Em 2009, houve bastante movimentação no mercado por parte da empresa, contando com a aquisição da Prontodente e da OdontoServ. Ainda em 2009, houve a constituição da *joint-venture*, no México. Foram iniciados os trabalhos na atuação da *Dental Partner*, assim como a empresa iniciou a associação com a Bradesco Dental.

Passando dois anos, em 2011, uma nova modalidade de planos iniciou sua comercialização: PME Bradesco. Este fato alavancou ainda mais as vendas de planos de saúde da empresa. Logo no ano seguinte, 2012, ainda em associação com a Bradesco, foi iniciada a

comercialização de planos individuais pelo banco, assim como, nesse mesmo ano, foi realizada a aquisição compartilhada da Papaiz.

No ano de 2013 se iniciou uma parceria com a BB Seguridade, e como fruto desse acontecimento, foi criada a Brasidental. Com o crescimento da parceria com a Bradesco e o vislumbre da nova parceria com o Banco do Brasil, uma reorganização societária fez com que a Bradesco fosse alçada ao posto de controlador majoritário.

Com o avanço tecnológico no mercado da saúde, surgiu a necessidade de a empresa lançar o seu aplicativo: Rede UNNA. Essa aplicação serve para conectar os mais de 28 mil credenciados da companhia. Em 2016, deu-se início a parceria com a A.C.Camargo Câncer, o Instituto de Educação Boa Vontade e Instituto Cejam.

Ainda no intuito de adaptar seu mercado ao avanço tecnológico, a Odontoprev criou o seu *e-commerce*, para vender seus produtos de forma *online* e digital, abrangendo ainda mais clientes e aumentando sua capilaridade.

Em 2018 ocorreu uma das grandes aquisições da companhia: o Grupo Odonto *System* foi incorporado a Odontoprev, trazendo consigo um aumento exponencial do número de clientes na região Nordeste, onde o grupo Odonto *System* mantinha um mercado expressivo.

Nos anos seguintes, houve a criação do Dentista *Online*, assim como a aquisição da Mogidonto Planos Odontológicos e da Boutique Dental. Até o fim do ano de 2022, o grupo Odontoprev possuía mais de oito milhões de clientes, fato que o credencia como um dos grandes grupos que oferece serviço odontológico na América Latina.

2.2 Business intelligence

BI é uma disciplina primordial para as organizações modernas, que permite coletar, analisar e interpretar dados para tomar decisões estratégicas (RAMOS; SANTOS, 2006). Envolve transformação de dados brutos em informações que são utilizadas para garantir eficiência em diversos aspectos, seja operacional, para identificar oportunidades de crescimento, avaliar desempenho e prever tendências.

Em resumo, o *BI* se refere ao processo de coletar, organizar, analisar e apresentar informações que sejam relevantes e possam dar apoio nas tomadas de decisões em uma organização. No geral, essas informações são obtidas através de dados internos ou externos, como registros de vendas, dados operacionais, dados de mercado, redes sociais e muitas outras opções.

Pode-se dizer que o processo de *BI* envolve algumas etapas: coleta de dados, utilizando fontes diversas, como banco de dados, sistemas de gestão empresarial, aplicativos, fontes de dados externas; transformação de dados, onde os dados brutos que foram coletados são transformados para estarem em um formato mais útil, passando por processos como limpeza e integração, por exemplo; armazenamento de dados, dentro de repositórios de dados, como *data Warehouse* ou *data lakes*, para facilitar o acesso e a consulta; análise de dados, onde *insights* são descobertos, tendências são identificadas e métricas são criadas para avaliação de desempenho, dentre outros marcadores para tomada de decisão; apresentação dos dados, de forma simples e acessível, por meio de painéis interativos, relatórios e gráficos (RAMOS; SANTOS, 2006).

Segundo Boulicat *et al.* (2004), existem uma gama de aplicações possíveis para o *Business Intelligence*, em diversas áreas de negócios, podendo ser citadas algumas:

a) Análise de vendas, monitorando o desempenho das vendas, identificando tendências de mercado ou previsão de demandas;

b) Gestão de desempenho corporativo, facilitando o acompanhamento do desempenho de uma corporação em relação as metas e objetivos; análise de custos e eficiência operacional, auxilia na identificação de oportunidades de economia de custos e melhor eficiência;

c) Gestão de recursos humanos, através de acompanhamentos de métricas de RH, como rotatividade de funcionários e níveis de satisfação;

d) Tomadas de decisões estratégicas, pelo fornecimento de dados para planejamento estratégico, auxiliando na avaliação da concorrência e na identificação de oportunidades de crescimento.

Existem algumas ferramentas e tecnologias que são utilizadas no apoio a *Business Intelligence*. Podem ser citadas algumas como as mais usuais: ferramentas de relatórios e *dashboards*, como *Tableau*, *Power BI*, *QlikView*, que permitem a criação de relatórios interativos e painéis de controle; ferramentas de extração, transformação e carga (ETL), utilizadas para coletar, limpar e transformar dados antes de armazená-los; bancos de dados empresariais, onde *data warehouses*, *data lakes* e bancos de dados relacionais podem ser utilizados para armazenar dados; linguagens de consulta, sendo o *SQL* a linguagem mais utilizada para consultar dados armazenados (MACEDO, 2022).

Figura 1 - Etapas que compõem o Business Intelligence de forma detalhada



Fonte: Hashtag (2023).

2.3 Power BI

O *Power BI* é uma suíte de ferramentas de análise de negócios desenvolvida pela Microsoft que permite que você visualize seus dados e compartilhe insights em toda a organização, podendo os incorporar em um aplicativo ou site. Ele é projetado para usuários de negócios que desejam criar relatórios e painéis interativos sem depender exclusivamente de equipes de TI ou especialistas em dados (FERRARI; RUSSO, 2017).

Existem alguns conceitos e componentes do *Power BI* que são fundamentais para seu entendimento como ferramenta (CLARK, 2017):

a) *Power Bi Desktop*, que é uma aplicação para Windows que permite criar relatórios e painéis interativos, oferecendo uma interface gráfica amigável para importar dados de várias fontes, realizando transformações nos dados e criando visualizações;

b) Os conjuntos de dados, que são coleções de dados que podem ser importados ou conectados ao *Power BI* para criação de relatórios, podendo ser originados de uma variedade de fontes, como bancos de dados, planilhas em Excel, serviços *online*, dentre outros;

c) As transformações de dados, que permite que o usuário limpe e modele os dados antes de trabalhar nos visuais, através de diversas ferramentas de ETL (*Extract, Transform, Load*), usadas para extrair, transformar e carregar os dados, como o próprio nome sugere;

d) As visualizações, através de gráficos, tabelas e outros vários elementos visuais que auxiliam na representação dos dados de forma mais compreensível; os relatórios, que são

páginas interativas que contém visualizações, gráficos e tabelas, onde os usuários podem criar relatórios dinâmicos para explorar dados de maneira mais aprofundada;

e) Os painéis, que são uma junção de visualizações de relatórios que podem ser compartilhadas e atualizadas em tempo real, e são ideais para apresentar de forma consolidada os principais indicadores de desempenho da análise;

f) *Power BI Service*, é uma plataforma *online* que permite que o usuário publique, compartilhe e colabore em relatórios e painéis, em tempo real, oferecendo recursos de atualização agendada, por exemplo;

g) *Gateway* de dados, que permite a conexão do *Power BI* com fontes de dados locais, garantindo assim a automação da atualização dos conjuntos de dados; as integrações com outras ferramentas Microsoft, como Excel, por exemplo;

h) *Power Query* e *Power Pivot*, que são ferramentas integradas no *Power BI* que auxiliam na transformação de dados e na criação de modelos de dados avançados;

i) *Power BI Mobile*, que permite o acesso a relatórios e painéis via dispositivos móveis.

O *Power BI* é uma ferramenta muito poderosa para transformação de dados brutos em informações relevantes, permitindo que quem a utilize seja capaz de realizar tomadas de decisões estratégicas com embasamento em análises visuais e insights de dados.

2.4 Orange Data Mining

Orange é uma estrutura abrangente baseada em componentes para aprendizado de máquina e mineração de dados. Destina-se tanto a usuários experientes quanto a pesquisadores em aprendizado de máquina que desejam escrever *scripts Python* para prototipar novos algoritmos enquanto reutiliza o máximo de código possível, e para aqueles que estão apenas inserindo o campo que pode desfrutar do ambiente de programação visual poderoso e fácil de usar (DEMSAR; ZUPAM, 2013).

Ainda segundo os autores o *Orange* suporta várias tarefas que vão desde o processamento de dados, até modelagem e avaliação, como exemplos tem-se:

- a) Gerenciamento e pré-processamento de dados, como amostragem, filtragem, escalonamento, discretização, construção de novos atributos e similares;
- b) Indução de modelos de classificação e regressão, incluindo árvores, classificador Bayesiano, abordagens baseadas em instâncias, regressão linear e logística e apresenta suporte para vetores;

- c) Ajuste de previsão de probabilidades de modelos de classificação; métodos descritivos, como regras de associação e agrupamento;
- d) Métodos para avaliação e pontuação de modelos de previsão, incluindo diferentes esquemas de resistência e variedades de método de pontuação e abordagens de visualização.

2.5 Problemas de regressão e classificação

Nesse tópico, buscar-se-á entender os dois principais tipos de problemas abordados pelo *Orange Data Mining*, para se poder integrar o contexto do problema avaliado nesse trabalho.

2.5.1 Regressão

Em um problema de regressão, é prever um valor numérico ou contínuo com base em um conjunto de características de entrada, ou seja, tentamos estimar uma relação funcional entre as variáveis de entrada e as de saída. Para esse tipo de problema, podemos definir as variáveis de saída como contínuas e quantitativas (CHEIN, 2019). Prever preços de mercado ou valores de imóveis seriam bons exemplos de possíveis variáveis de saída para um problema de regressão.

Existem alguns métodos utilizados para avaliar as saídas em um modelo de regressão, como o Erro Quadrático Médio (MSE), Erro Absoluto Médio (MAE), Coeficiente de Determinação (R^2), entre outras métricas. O MSE (Equação 1) é comumente utilizado para verificar a acurácia dos modelos, e ele dá maior peso aos maiores erros, já que, no seu cálculo, cada erro é elevado ao quadrado, e após isso é calculada as médias desses erros quadráticos.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2. \quad (1)$$

Já o MAE (Equação 2) nos diz a média da diferença absoluta entre os valores observados e os valores preditos pelo modelo. Por não apresentar o termo ao quadrado, o MAE não dá tanto peso aos maiores erros.

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n}. \quad (2)$$

O coeficiente de determinação (R^2) é uma medida estatística que avalia a qualidade de um ajuste de um modelo de regressão linear (Equação 3). Essa métrica varia entre 0 e 1. Quando o coeficiente está próximo de 0, significa que o modelo não representa uma explicação dos dados analisados, enquanto que, quando o coeficiente se aproxima de 1, o modelo avaliado explica muito bem as variações dos dados do problema. O R^2 é uma medida de fácil interpretação, por isso se torna vantajoso para utilizar na análise.

$$R^2 = 1 - \frac{RSS}{TSS}, \quad (3)$$

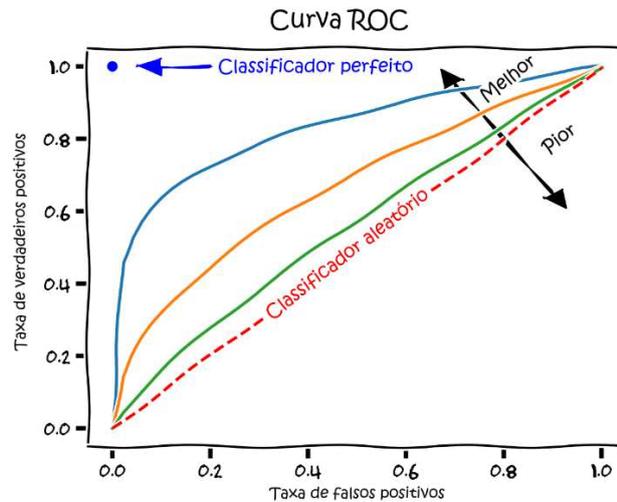
onde RSS é a soma dos quadrados dos resíduos e TSS é a soma dos quadrados das diferenças das observações para as médias.

2.5.2 Classificação

Um problema de classificação tem como objetivo atribuir uma categoria ou classe a algo com base nos dados de entrada do problema. Nesse caso, não há uma previsão de um valor contínuo, mas sim uma atribuição de uma classificação discreta. Nesse tipo de problema, a variável de saída é categórica e discreta (SILVEIRA: BULLOCK, 2017). Por exemplo, para definir qual animal está contido em uma imagem, precisa-se categorizar as imagens pelos animais (gato, cachorro, cavalo, etc), ou seja, rotular os dados em categorias.

Segundo Rodrigues (2019), existem alguns métodos para avaliar o desempenho de um modelo de classificação. Pode-se citar a Precisão, Revocação, F1-Score, Matriz de Confusão e Área sob a curva ROC (descrita na imagem):

Gráfico 1 – Área sob a curva ROC



Fonte: Lean Saúde (2022).

Verifica-se na imagem acima que o eixo x define a taxa de falsos positivos, ou seja, relativo à testes realizados que geram resultados positivos, porém não estão corretos; o eixo y mostra a taxa de verdadeiros positivos, que são justamente os casos em que o positivo se mostra correto. Quanto maiores os valores de x e y, melhor será o ajuste do modelo testado.

No método de Precisão (Equação 4), avalia-se a proporção de previsões corretas feitas pelo modelo, se comparado ao total de previsões positivas.

$$\text{Precisão} = \frac{VP}{VP+FP}, \quad (4)$$

onde VP são os verdadeiros positivos, ou seja, as previsões corretas, e FP são os falsos positivos, que são as previsões erradas.

Na Revocação, avalia-se a proporção de verdadeiros positivos dentre todas as avaliações corretas (que seria a soma de verdadeiros positivos e falsos negativos).

$$\text{Revocação} = \frac{VP}{VP+FN}, \quad (5)$$

onde FN são os falsos negativos.

No método F1-Score é realizada uma média harmônica entre os dois métodos citados anteriormente (Precisão e Revocação).

$$\text{F1 - Score} = \frac{2(\text{Precisão})(\text{Revocação})}{\text{Precisão}+\text{Revocação}}. \quad (6)$$

A Matriz de Confusão é uma forma gráfica de visualizar as previsões corretas ou incorretas do modelo, em forma de tabela. A Figura 2 expõe um exemplo de matriz.

Figura 2 - Exemplo de Matriz de Confusão

		Detectada	
		Sim	Não
Real	Sim	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte: Rodrigues (2019).

Na matriz, vê-se quatro campos: Verdadeiro Positivo (VP), que indica a quantidade de observações corretas do modelo; Falso Negativo (FN), ou Erro do tipo II, que indica os erros em que o modelo previu classe negativa mas na verdade era positiva; Falso Positivo (FP), ou Erro do tipo I, que seria os casos dos erros onde o modelo previu a classe positiva, porém na realidade pertence a classe negativa; e têm-se os Verdadeiros Negativos (VN), onde o modelo previu corretamente a classe negativa.

A Área Sob a Curva ROC (AUC-ROC) é uma medida que diz o quanto um modelo é capaz de distinguir entre as classes. Quanto maior essa medida, melhor será a capacidade do modelo de classificar corretamente os dados (SANTOS, 2019). Existe um gama de algoritmos que modelam problemas de classificação. Alguns deles são Regressão Logística, Máquinas de Vetores de Suporte (SVM), Árvores de Decisão e Florestas Aleatórias, Redes Neurais, etc.

De forma resumida, a principal diferença entre os problemas de regressão e de classificação consiste no tipo de variável de saída: se for uma variável contínua, lida-se com um problema de regressão; caso seja discreta, estar-se diante de um problema de classificação.

2.6 Modelos e métodos avaliados

Nesse tópico, serão avaliados todos os modelos e métodos utilizados na metodologia da situação problema colocada em pauta.

2.6.1 *Random Forest*

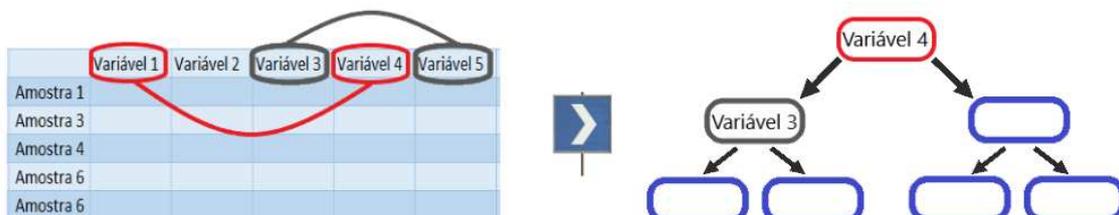
Random Forest é um algoritmo que, como o próprio nome fala, funciona como uma “floresta aleatória”. Ele funciona criando muitas árvores de decisão (algoritmos onde vários pontos, chamados de nós, são utilizados para definir qual caminho o algoritmo irá seguir, fluindo para os seus “ramos”), de maneira aleatória. Para definir o resultado final, ocorre uma espécie de votação, onde cada árvore será um candidato para a finalidade (ZHANG; MA, 2012).

Para melhor entender como funciona o algoritmo, é necessário conhecer os métodos *ensemble*, dos quais ele faz parte. De acordo com Didatica (2023), nesse método ocorre uma resposta agregada entre vários modelos testados. Geralmente são modelos mais robustos e operacionalmente eficientes, apesar de computacionalmente mais complexos.

Como o modelo *Random Forest* é definido como uma agregação de várias árvores de decisão, é criada uma estrutura semelhante a um fluxograma, onde cada um dos “nós” possui uma verificação de decisão, e, a depender do que for escolhido, o fluxo segue para o ramo subsequente, até que se chegue no último ramo da árvore (DIDATICA, 2023).

No *Random Forest*, a definição da variável não acontece tomando como base todas as variáveis disponíveis, mas sim escolhendo duas ou mais variáveis, de maneira aleatória, e então realiza os cálculos com as amostras selecionadas, para então definir qual variável estará no primeiro nó. Para a escolha do próximo nó, novamente serão escolhidas duas ou mais variáveis para compor a amostra, porém dessa vez não serão selecionadas as variáveis já escolhidas anteriormente. Seguindo esse padrão, a árvore será construída até seu último nó. Na figura, temos um resumo da construção citada (CUTLER *et al.*, 2012).

Figura 3 - Resumo das escolhas das variáveis para compor os nós de uma árvore aleatória



Fonte: Didatica (2023).

Para a construção das próximas árvores, é necessária a repetição do processo citado anteriormente. Como o processo ocorre de forma aleatória, mesmo seguindo o mesmo passo a

passo, muito provavelmente serão construídas árvores diferentes entre si. Quanto mais árvores criadas, melhor será a acurácia do modelo. Se o problema estudado for de regressão, o resultado final será a média dos resultados previstos. Já, se o problema for de classificação, o resultado considerado será o que mais vezes se repetir.

2.6.2 Regressão Linear (*Linear Regression*)

A regressão linear é um modelo estatístico muito utilizado para resumir dados. Também tem sua utilização atrelada à modelagem da relação entre uma variável resposta à uma ou mais variáveis independentes (também chamados preditores). Com a utilização dessa ferramenta, deseja-se prever comportamentos de variáveis de acordo com as observações feitas, as variáveis disponíveis.

Na regressão linear, busca-se encontrar uma equação linear que descreva da melhor forma as observações feitas, ou seja, que melhor represente o conjunto de dados disponíveis (CHEIN, 2019). Geralmente, essa equação é definida da seguinte forma:

$$Y = aX + b, \tag{7}$$

onde Y é a variável que se está buscando prever; X é o preditor; a é o coeficiente angular, que representa a inclinação da reta de regressão; e b é o intercepto (ou coeficiente linear), que indica o ponto onde a reta de regressão cruza o eixo vertical (X = 0) (COELHO *et al.*, 2007).

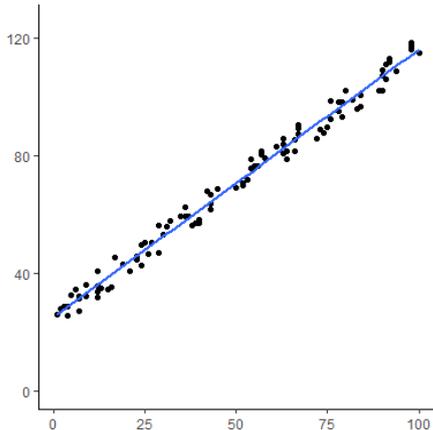
Com todas as variáveis explicadas, o objetivo do modelo de regressão é encontrar valores de a e b ótimos, ou seja, de forma que os erros sejam mínimos (a diferença entre os valores observados e os valores preditos seja a menor possível, dadas as informações do problema).

O método mais comumente utilizado para atingir a otimização do modelo de regressão é chamado de método dos mínimos quadrados. O principal objetivo desse método é encontrar a melhor linha (ou reta) que minimize a soma dos quadrados das diferenças (Equação 8) entre os valores observados e os valores preditos (ALMEIDA, 2015).

$$\sum(Y_i - \hat{Y}_i)^2 \tag{8}$$

Após definido o modelo de regressão através do método de mínimos quadrados, geralmente se desenha a linha no gráfico onde os dados estão dispostos, com o intuito de visualizar o comportamento da reta diante dos dados do problema. O Gráfico 1 exemplifica o traçado da reta de regressão linear.

Gráfico 2 - Reta de regressão linear aplicada sobre dados



Fonte: elaboração própria.

Dentro da ferramenta que será utilizada (*Orange*), a regressão linear se utiliza do método dos mínimos quadrados na sua análise.

2.6.3 Neural Network (Rede Neural)

As redes neurais têm esse nome devido a sua inspiração no funcionamento do cérebro humano, dada semelhança entre as funções de ambos. Tais redes são usadas com bastante frequência para lidar com problemas computacionais, processamento de linguagem natural, reconhecimento de padrões, etc.

Redes neurais são modelos matemáticos formulados e compostos por camadas, que chamamos de “neurônios” (vide analogia ao funcionamento do cérebro humano), que são conectados entre si. Cada um desses neurônios funciona como uma unidade de processamento que recebe sinais de entrada, realiza uma “tradução” e emite as saídas (como uma sinapse cerebral) (GOLDBERG, 2016). Cada conexão entre neurônios possui um peso específico, que identifica o quanto um influencia no fluxo do outro. O Quadro 1 expõe alguns tipos de neurônios, suas características, vantagens e desvantagens:

Quadro 1- Tipos de neurônios, suas características, vantagens e desvantagens

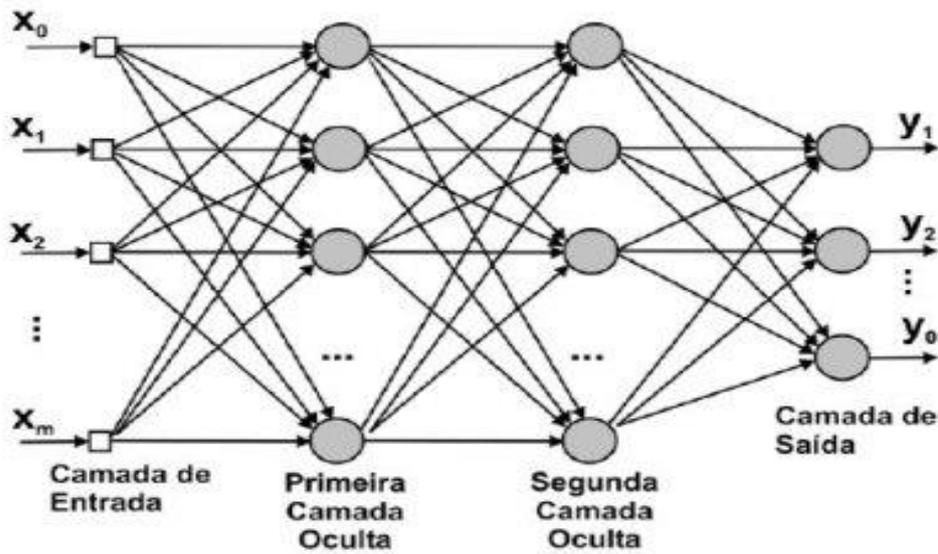
Tipo de neurônio	Principais Características	Desvantagens	Vantagens
Perceptron	Usa variáveis binárias de entrada	Não possui camada oculta	Sistema de base para demais tipos de neurônios
	Introduz pesos nas variáveis de entrada		
	Produz uma única saída binária	Aprendizado de recursos pouco complexos	Utilização para dados separados linearmente
Perceptron Multicamadas	Possui camadas ocultas empilhadas	Grande consumo de memória para aprendizado	Camadas ocultas com pesos distintos permite o aprendizado de recursos mais complexos
		Função de ativação Step (linear e não suave)	
Sigmóide	Utiliza backpropagation	Redes de neurônios mais complexos	Ajuste repetidos dos pesos minimiza a diferença entre a saída real e a saída desejada
	Função de ativação Sigmóide		Função de ativação estável e suave Permite classificação de limites ou padrões de decisões não lineares

Fonte: Goldberg (2016).

As redes neurais são treinadas através de um processo de aprendizado. Funciona da seguinte forma: um conjunto de dados é alimentado para a rede, que gera os valores previstos pelo modelo. A seguir, são calculados os erros, ou seja, todas as diferenças entre os valores previstos pelo modelo e os valores observados (os dados). Esse erro é realocado a rede, e essa etapa ajusta os pesos entre os neurônios do modelo. Tal processo se repete várias vezes até que o modelo alcance uma eficiência aceitável (depende do objetivo) (KOVÁCS, 2006).

Existem alguns tipos de redes neurais que merecem destaque, como Redes Neurais Feedforward (FNN) (Figura 4), que têm direção única, do início ao fim, e não apresentam ciclos; Redes Neurais Convolucionais (CNN), que são projetadas para dados de grade (como imagens), identificando padrões em diferentes regiões; Redes Neurais Recorrentes (RNN), usadas para processar sequência de dados, mantendo-se o estado interno da rede e atualizando ao longo do tempo; Redes Neurais Generativas (GANs), onde duas redes neurais concorrentes (uma geradora e uma discriminadora) são usadas para gerar dados que sejam realistas.

Figura 4 - Imagem que exemplifica uma Rede Neural Feedforward (FNN)



Fonte: Facure (2023).

Existem várias aplicações para o modelo de redes neurais, como reconhecimento de imagens, processamento de linguagem natural (em traduções automáticas, *chatbots*), automatização de tarefas, previsões, classificações, jogos, robótica, etc.

2.6.4 KNN (*K-vizinhos mais próximos*)

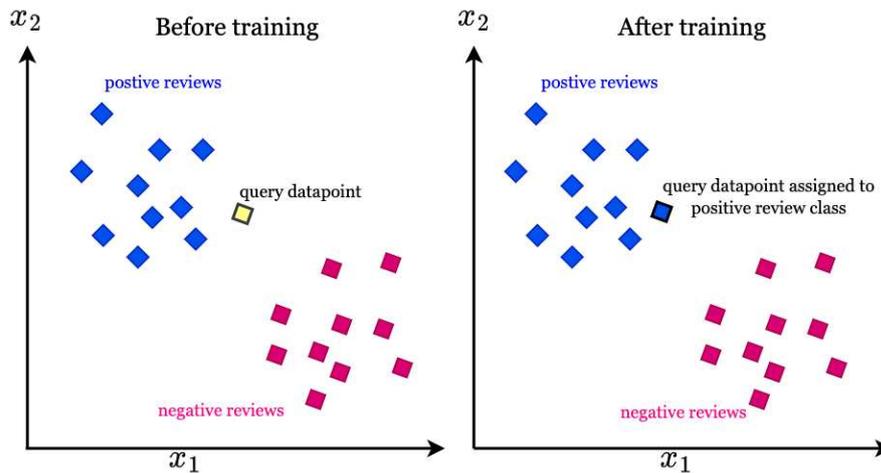
Esse algoritmo é uma técnica de aprendizado de máquina que se enquadra na categoria de classificação supervisionada, onde existe o acompanhamento de uma pessoa especialista no tema que se enquadra a situação problema. É um método simples e intuitivo para problemas de classificação, sendo muito utilizado por eficácia e por se mostrar muito volátil, sendo aplicado em diversas situações reais (GUO *et al.*, 2003).

Tal algoritmo funciona agrupando objetos semelhantes, levando em consideração determinadas características. O funcionamento dele leva em conta uma série de etapas para que haja sucesso em sua análise.

De acordo com Ferrero (2009), tais etapas são coleta de dados para treinamento, escolha do parâmetro k , que vai ser um número inteiro positivo. Esse parâmetro vai representar a quantidade de vizinhos mais próximos, ou seja, o número de objetos próximos ao objeto em questão. Com os dados de treinamento já definidos e o treinamento realizado, pode-se classificar um novo objeto na análise: avaliando tal objeto, o algoritmo procura os k objetos mais próximos dele que fazem parte do conjunto de treinamento, baseando-se em medidas de distância (distância euclidiana, por exemplo). Os k -vizinhos mais próximos desse objeto que

não estão no conjunto de treinamento serão classificados de acordo com as características desse objeto escolhido inicialmente (CAMBRONERO; MORENO, 2006). O Gráfico 2 exemplifica o funcionamento do algoritmo.

Gráfico 3 - Classificação de objeto antes/depois do treinamento dos dados pelo algoritmo KNN



Fonte: Intuitive (2023).

Tal algoritmo tem algumas aplicações dentro de cenários de relevância, como classificação de documentos, recomendação de produtos, diagnósticos médicos, detecção de fraudes, visões computacionais.

Apesar de ser um algoritmo amplamente utilizado em diversos aspectos, ele apresenta uma série de limitações, segundo Cambronero e Moreno (2006), que devem ser levadas em consideração no momento da sua utilização:

- a) É sensível a Outliers, ou seja, seu desempenho pode ser afetado por valores que se mostram fora da curva dentre os dados avaliados;
- b) Para realizar o treinamento, o algoritmo exige que os dados de treinamento sejam representativos e tenham um volume considerável;
- c) Existe o desafio da escolha do parâmetro k , que afeta diretamente no funcionamento do algoritmo e na sua eficácia;
- d) Existe um grande custo computacional, principalmente quando se trata dos cálculos das distâncias entre os objetos, especialmente se houver um grande volume de dados.

2.6.5 Cross-Validation (Validação Cruzada)

É uma técnica de aprendizado de máquina e estatística, utilizada para medir o desempenho de um modelo no que se refere a generalização, e também apresenta papel preponderante no combate a situações de *overfitting*, que seria a situação em que um modelo faz um ajuste exagerado dos dados, gerando resultados que se apresentam fora da realidade do contexto do problema (CUNHA, 2019).

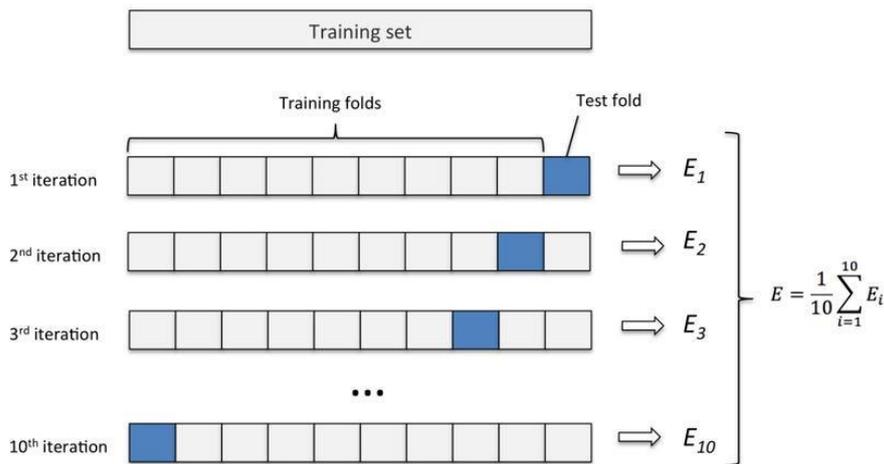
Essa técnica funciona da seguinte maneira: o conjunto de dados analisado é dividido em duas partes, uma delas chamada de conjunto de treinamento (funciona para treinar o modelo), e a outra, conjunto de teste (usado para testar o desempenho do modelo testado). Nesse caso, ao invés de dividir os dados em uma única divisão de treino e teste, esse modelo alterna tais conjuntos, fazendo com que, a cada iteração, os dados de treino e teste sejam diferentes.

Cunha (2019) detalha as três grandes etapas na construção da validação cruzada, que também pode ser observada na Figura 5:

- a) A divisão dos dados em k subconjuntos de tamanhos iguais (ou aproximados), onde cada uma dessas partições é chamada de “*fold*”;
- b) O modelo é treinado k vezes, e em cada iteração, uma das partições é utilizada como teste, enquanto as outras é usada como conjunto de treinamento, havendo variação entre quais partições são usadas para treinamento e qual é usada para teste;
- c) Ao fim das iterações, é feita uma avaliação agregada de desempenho, onde algumas métricas de erro são agrupadas para formar uma avaliação final de desempenho do modelo.

O autor ainda cita as vantagens do método, a saber: quando se compara com as técnicas em que somente há uma divisão possível de treino e teste, nota-se que a validação cruzada oferece uma opção de avaliação de desempenho mais robusta, pois oferece a capacidade de variar as divisões de treino e teste; tal técnica auxilia na identificação de *overfitting* do modelo, avaliando se há uma grande diferença no conjunto de teste para cada iteração e mudança da partição avaliada; por realizar o treino e o teste diversas vezes, permite que o a quantidade de dados utilizada em cada conjunto (treino e teste) seja maior do que habitualmente seria; o desempenho avaliado do modelo se mostra mais preciso, pois a métrica utilizada para avaliar o desempenho se mostra como uma média das métricas de cada iteração.

Figura 5 - Funcionamento da técnica de Validação cruzada



Fonte: Medium (2023).

Existem algumas técnicas de validação cruzada com especificidades que podem apresentar utilidade a depender do contexto ou problema em questão. Pode-se citar a Validação Cruzada “leave-one-out” (LOOCV), Validação Cruzada Estratificada, Validação Cruzada por grupos, etc.

3 APLICAÇÃO METODOLÓGICA

Para se chegar à solução do problema colocado em pauta, utilizou-se do auxílio das duas ferramentas apresentadas nos tópicos anteriores: O *Power BI*, que será utilizado para realizar a análise gráfica das variáveis em análise, e o *Orange Data Mining*, que servirá como aporte estatístico para realizar testes e comparar as variáveis, traçar análises mais profundas dos modelos e entender qual ou quais melhores se ajustam à realidade dos dados.

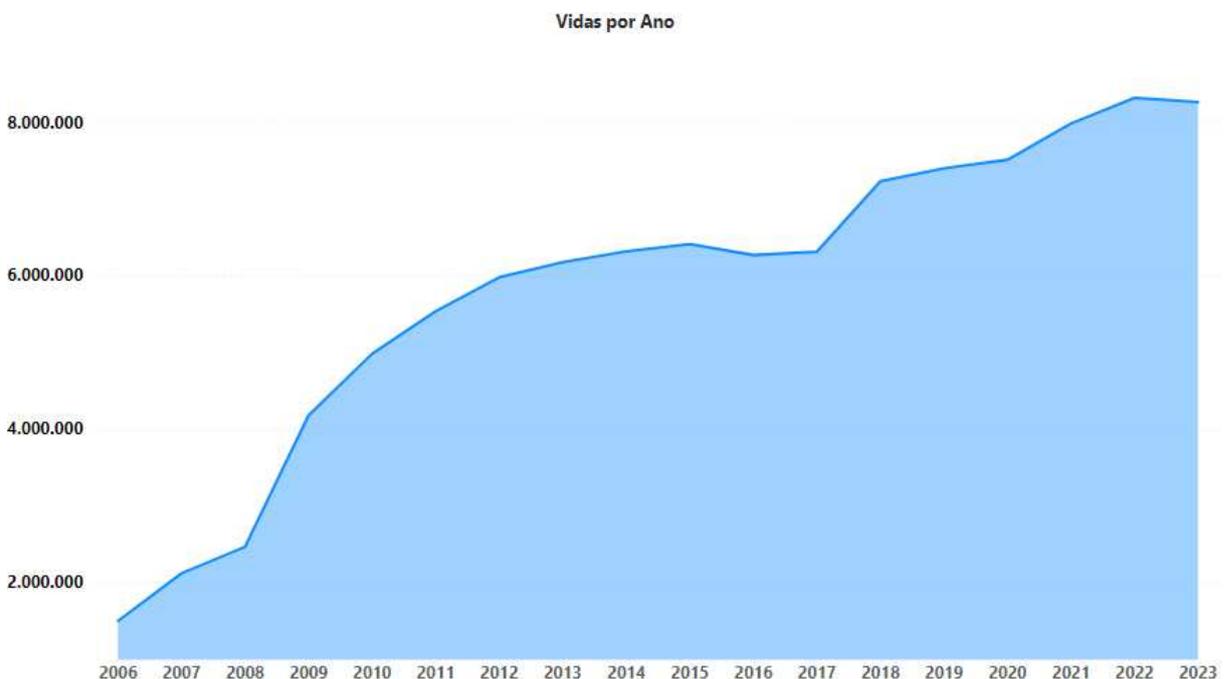
3.1 Variáveis utilizadas

Para a análise do problema em questão, foram avaliadas 10 variáveis: os casos de Covid-19, assim como a quantidade de óbitos por covid-19, ambas avaliadas do período de 26 de fevereiro de 2020 até 19 de agosto de 2023; a receita líquida, o ticket médio, a quantidade de vidas, a sinistralidade, EBTIDA ajustado, margem EBTIDA, lucro líquido e margem líquida, todas referentes ao grupo Odontoprev, entre os anos de 2006 e 2023 (até o primeiro trimestre). Todas as variáveis avaliadas no presente trabalho são anuais, ou seja, os valores são avaliados ano a ano.

3.2 Análise no *Power BI*

O Gráfico 4 se refere a evolução da quantidade de vidas do grupo Odontoprev. Avaliando graficamente, pode-se concluir que houve uma evolução crescente no número de vidas do grupo ao passar dos anos, o que indica crescimento da marca e pode, juntamente com outros indicadores, justificar a relevância da empresa dentro do mercado odontológico. Nos anos referentes à pandemia, notamos que a quantidade de vidas se manteve em uma crescente similar a ocorrida em todos os anos da análise, o que não torna relevante a análise gráfica nesse caso.

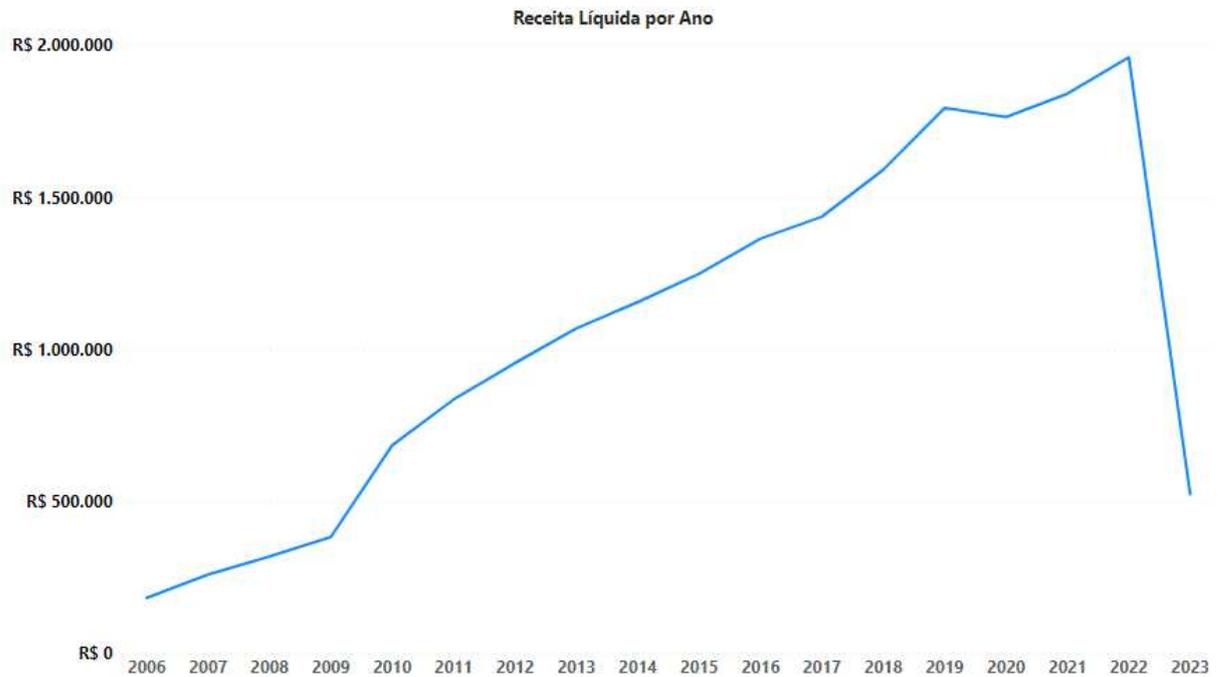
Gráfico 4 - Evolução da quantidade de vidas do grupo Odontoprev, por ano



Fonte: elaboração própria.

Outra análise gráfica muito importante que pode também indicar relevância da empresa no mercado é o comportamento da receita líquida do grupo (Gráfico 5).

Gráfico 5 - Evolução da receita líquida do grupo Odontoprev, por ano



Fonte: elaboração própria.

Pode-se perceber que houve uma crescente constante na variável ao longo dos anos, com exceção apenas no último ano da análise (2023). Porém, essa queda se justifica pela premissa dos dados: para o ano de 2023, há apenas informações do primeiro trimestre do ano, o que nos dá um corte na análise, tornado inviável a projeção do ano em questão.

Assim como a receita líquida, o lucro líquido é um indicador importante de relevância do grupo no mercado, e, assim como as duas variáveis anteriores, essa também se mostra crescente em seu comportamento, como mostra o Gráfico 6.

Gráfico 6 - Evolução do lucro líquido do grupo Odontoprev, por ano



Fonte: elaboração própria.

Vale ressaltar dois pontos: no ano de 2017 houve um comportamento anômalo da variável, que se justifica pela criação do *e-commerce* do grupo, alavancando as vendas dos produtos de forma *online*; no ano de 2023, percebe-se mais uma vez uma queda vertiginosa na variável, que se dá pela justificativa da premissa: apenas existem dados referentes ao primeiro trimestre do ano, o que diminui a quantidade da variável de forma inevitável.

Um indicador primordial para avaliar a relevância e o crescimento de uma empresa no mercado é a sinistralidade. Pode-se defini-la como a relação entre o custo dado por acionar o plano odontológico (chamado de sinistro) e o valor que a operadora recebe por isso (chamado de prêmio). Ou seja, quanto maior a sinistralidade de um plano, mais o beneficiário o utiliza e menos a empresa ganha por isso. No Gráfico 7, pode-se ver o comportamento da sinistralidade da Odontoprev ao longo dos anos.

Gráfico 7 - Evolução da sinistralidade do grupo Odontoprev, por ano

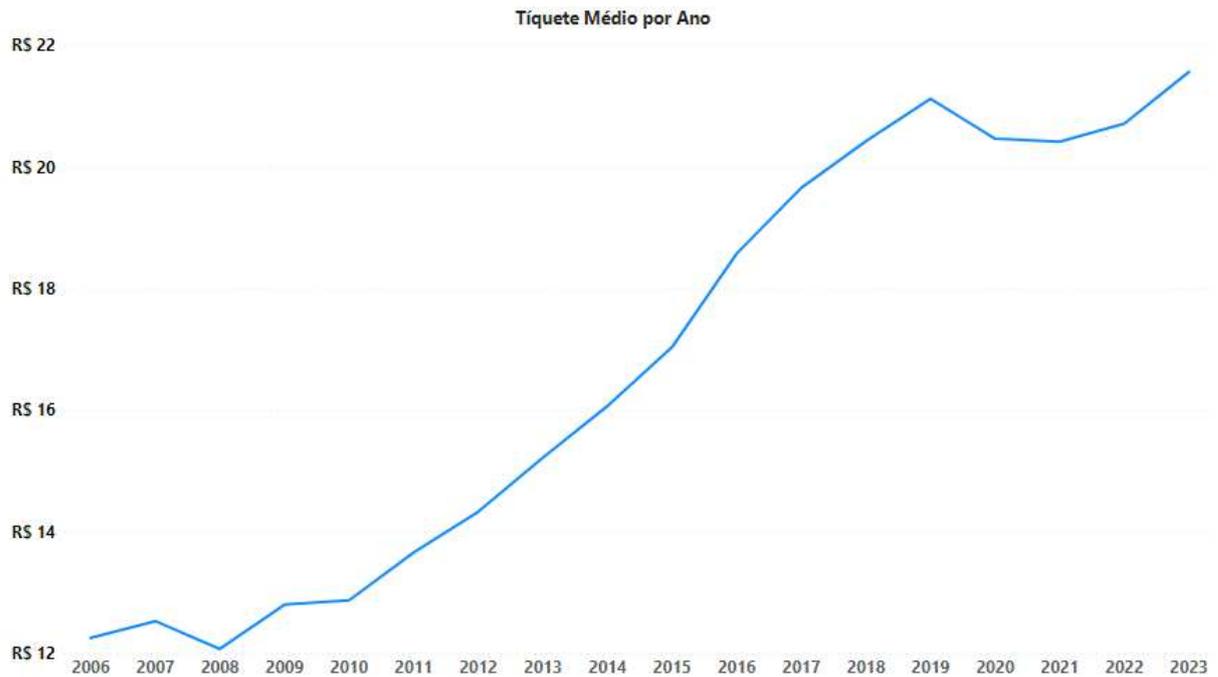


Fonte: elaboração própria.

No início do período, a sinistralidade estava entre 45% e 50%, e no fim, se mostrava abaixo de 35%. Essa queda é positiva para o grupo e demonstra, junto com outras variáveis, que a empresa, ao longo dos anos, conquistou equilíbrio entre a utilização dos beneficiários e o valor recebido pela mesma. Na prática, houve diminuição no número de sinistros e/ou aumento do valor recebido pela empresa por beneficiário.

Mais uma variável relevante é o ticket médio. Pode-se defini-la como a razão entre o faturamento da empresa e a quantidade de vendas. Ou seja, ela representa uma média do valor dos produtos vendidos pela empresa. Assim, se há um aumento no valor do ticket médio, pode-se inferir que o faturamento da empresa cresce, se houver a manutenção dos produtos vendidos pela mesma.

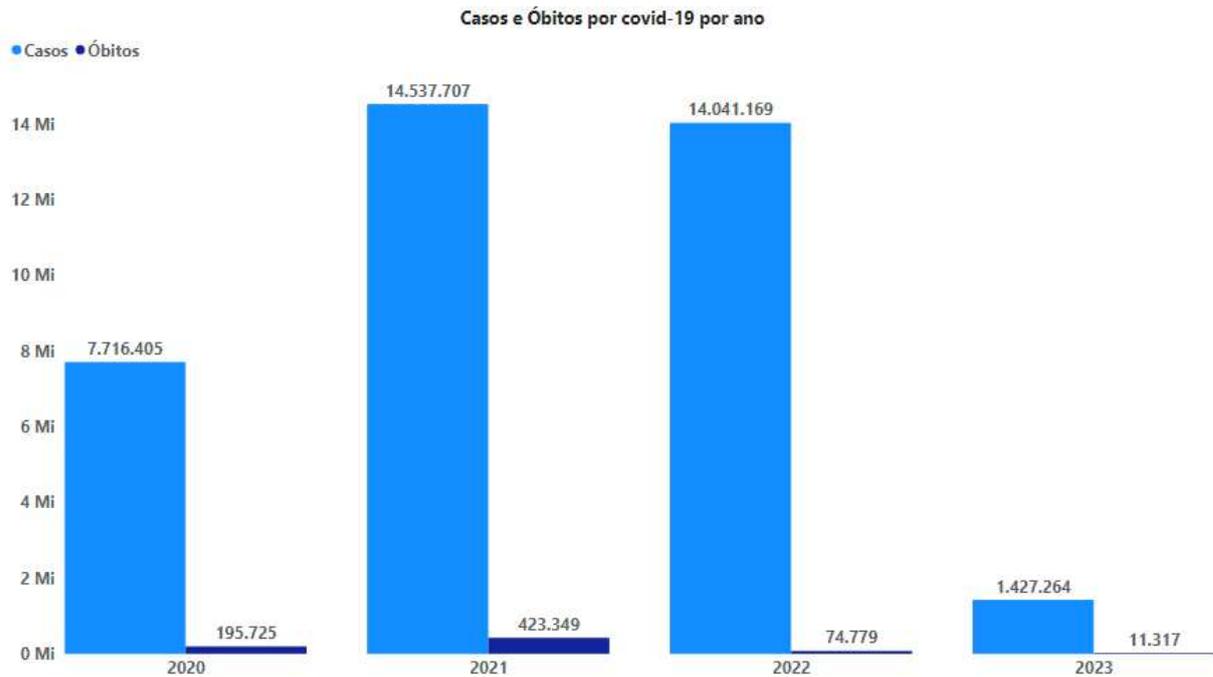
Gráfico 8 - Evolução do ticket médio do grupo Odontoprev, por ano



Fonte: elaboração própria.

Avaliando o Gráfico 8, pode-se notar que houve crescimento no valor da variável, havendo uma pequena queda no período referente à pandemia. Isso pode indicar uma pequena queda no faturamento no período citado. Porém, em 2022 e 2023, a variável retomou o crescimento, o que pode indicar um retorno do crescimento dos valores de faturamento, sendo devido ao retorno do estado de normalidade que foi mudado pela pandemia de Covid-19.

Gráfico 9 – Casos e óbitos por covid-19, por ano

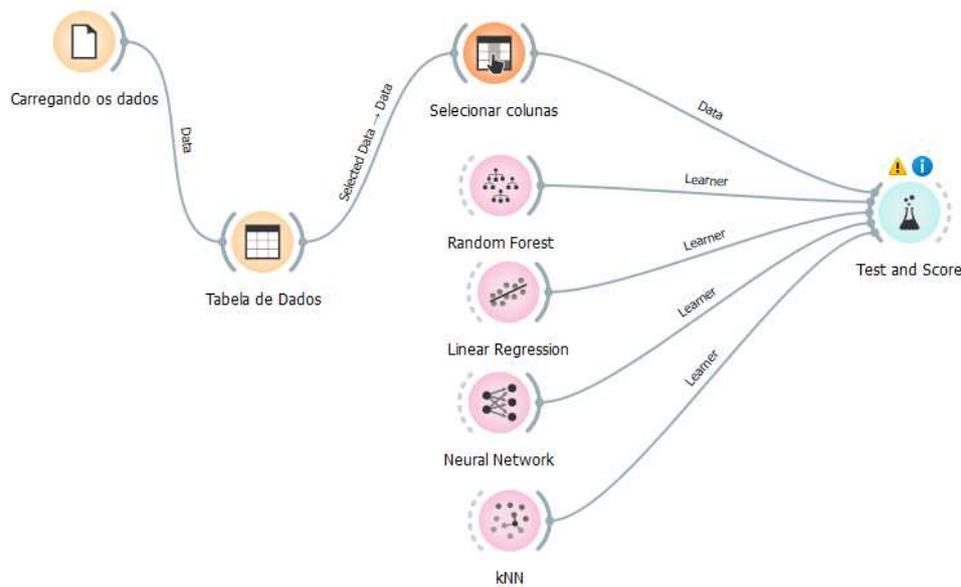


Fonte: elaboração própria.

3.3 Análise no *Orange Data Mining*

Para realizar a análise estatística, foi utilizado o *Orange Data Mining*, software que tem como sua principal função a análise e treinamento de dados, testando modelos de aprendizado de máquina, e que tem um visual simples e didático, com seu funcionamento vinculado a funções apresentadas em forma de blocos circulares em um quadro (Figura 6).

Figura 6 - Representação da tela de análises do *Orange Data Mining*



Fonte: Orange (2023).

Primeiramente, no quadro “Carregando os dados”, a tabela de dados, com as variáveis do problema, é carregada para o *Orange*, trazendo para a análise todas as informações necessárias para o objetivo do trabalho. No quadro “Tabela de dados”, cria-se uma visualização em formato de tabela, para que as variáveis do problema possam ser vistas, auxiliando na conferência da corretude dos dados carregados. Há uma etapa a mais sendo realizada antes de propriamente aplicar os modelos: a seleção de colunas. Essa etapa tem como função ajustar e selecionar as variáveis que realmente serão utilizadas na avaliação dos modelos, trazendo consigo melhorias significativas no processamento dos dados.

A etapa descrita como “*Test and Score*” tem como função avaliar o desempenho de cada um dos modelos utilizados para analisar os dados, através do método de validação cruzada, já descrito anteriormente. Foram quatro os modelos avaliados no problema: *Random Forest*, *Linear Regression*, *Neural Network* e KNN (todos descritos de forma detalhada na seção 2.6 do presente trabalho).

4. RESULTADOS E DISCUSSÃO

Pode-se avaliar os resultados, *a priori*, separadamente. Após isso, pode-se unir as avaliações e trazer uma conclusão mais completa sobre o estudo.

4.1 Resultados da análise no *Power BI*

Avaliando cada variável descrita na análise gráfica, pode-se inferir que, mesmo havendo um contexto de pandemia entre os anos de 2020 e 2023, o crescimento do grupo Odontoprev se manteve constante desde o início da divulgação dos dados avaliados, o que pode demonstrar que a empresa se mostra em um patamar de relevância dentro do mercado que a torna financeiramente sustentável, mesmo em um contexto de crise sanitária mundial.

O número de vidas, ou seja, de beneficiários do grupo no período pandêmico foi crescente. Porém, não foi um crescimento anômalo, já que durante todo o período analisado (entre 2006 e 2023), manteve esse comportamento. A receita líquida, o lucro líquido e a

sinistralidade mantiveram comportamento muito semelhante ao número de vidas, com crescimento constante no período completo, assim como no recorte do período referente à pandemia.

A única variável que apresentou comportamento diferente das demais foi o ticket médio. O ticket é a razão entre o faturamento do grupo e a quantidade de beneficiários, e existem cenários possíveis que justificam a queda desse índice no recorte pandêmico: a queda do faturamento ou o aumento vertiginoso no número de vidas. Mas, avaliando o gráfico de vidas citado anteriormente, nota-se que tal variável esteve em crescimento constante no período. Logo, a possibilidade que resta, foi que, no período, houve uma pequena queda no faturamento, porém, não refletindo na receita do grupo, que se manteve constante.

O que se pode inferir da análise gráfica é que, para o grupo Odontoprev, não houve impacto significativo no crescimento financeiro e da marca como um todo durante o período de crise sanitária mundial.

4.2 Resultados da análise no *Orange Data Mining*

A análise feita através do *Orange* é estatística, comparando modelos de regressão para entender qual desses apresenta melhor aderência aos dados do problema, e mostrando, a partir do melhor modelo, qual o grau de relação entre as variáveis, para entender se houve alguma influência relevante dos dados referentes a Covid-19 em relação às variáveis que avaliam o comportamento financeiro do grupo Odontoprev.

Quatro modelos foram testados para os dados do problema (*Random Forest*, *Linear Regression*, *Neural Network* e *KNN*), sendo utilizado o método de validação cruzada (*Cross Validation*) para realização dos testes para cada modelo. A medida utilizada para identificar qual modelo melhor se adequa aos dados foi o coeficiente de determinação, também conhecido como R^2 , coeficiente que auxilia na determinação da qualidade de ajuste de um modelo. O valor do tal coeficiente varia entre 0 e 1, sendo 0 quando o modelo não apresenta ajuste algum aos dados, e 1 quando o modelo se ajusta completamente os dados. Disto isto, verifica-se na Tabela 1 o valor da medida para cada um dos quatro modelos.

Tabela 1 - Tabela de valores de R^2 para cada modelo

Modelo	R ²
Random Forest	0.762
Linear Regression	0.704
Neural Network	-2.442
kNN	0.425

Fonte: elaboração própria

Observando os valores, a partir da premissa de que o melhor ajuste está mais próximo de 1, pode-se verificar que o modelo que melhor se ajusta aos dados é o *Random Forest*, com o valor do coeficiente igual a 0,762. Analisando cada um dos valores, verifica-se que a *Linear Regression* apresentou valor bem semelhante ao *Random Forest* (0,704), e que o KNN teve um resultado bem menos satisfatório (0,425). Nota-se que a *Neural Network* apresentou valor negativo para o coeficiente, o que não se aplica, visto que existe a necessidade de o valor estar no intervalo entre 0 e 1. Dessa forma, pode-se retirar tal modelo da análise.

Observando o valor apresentado pelo melhor modelo, ou seja, o valor de 0,762 obtido pelo *Random Forest*, pode-se inferir que existe aproximadamente 76,2% de ajuste dos dados, ou seja, a relação entre os dados relativos as variáveis referentes a Covid-19 e aos valores das variáveis que trazem os números de faturamento do grupo Odontoprev é de 76,2% aproximadamente. Em resumo, as variáveis relativas a Covid-19 explicam 76,2% dos números relativos ao faturamento do grupo.

5. CONSIDERAÇÕES FINAIS

A partir da análise gráfica realizada utilizando o *Power BI*, nota-se que os valores das variáveis relativas ao crescimento financeiro e, conseqüentemente, como marca, do grupo Odontoprev, se mantiveram em crescimento constante, mesmo quando se avalia o corte temporal relativo à pandemia. Apenas uma variável se mostra contrária a essa inferência: *ticket* médio. Porém, quando se avalia especificamente a forma como ela é calculada (razão entre faturamento e quantidade de beneficiários), o que pode ter causado essa pequena queda do *ticket* no corte temporal tem relação com uma possível queda no faturamento. Porém, como pode-se observar nas demais variáveis, essa queda não representou diminuição no cenário financeiro geral da marca, visto que todas as outras variáveis estão em crescente, seja no corte temporal pandêmico ou mesmo no período completo.

A análise gráfica não se mostra totalmente conclusiva somente quanto a interferência direta do período de pandemia de Covid-19 em relação ao crescimento financeiro e como marca do grupo Odontoprev. Portanto, foi realizada uma análise de relação entre as variáveis do problema. Dessa forma, utilizando o *Orange Data Mining*, foi realizado o teste com quatro modelos, escolhido o que apresentou o melhor ajuste em relação aos dados do problema, e, avaliando o coeficiente de correlação, viu-se que as variáveis relativas a covid-19 explicam cerca de 76,2% dos dados referentes as variáveis financeiras do grupo Odontoprev. De forma resumida, existe uma relação direta de 76,2% aproximadamente entre os números relativos a pandemia e o crescimento da empresa no período entre 2020 e 2023 (até o fim do primeiro trimestre).

Dessa forma, dada a relevância do grupo Odontoprev no mercado odontológico brasileiro e a relação direta entre seu crescimento financeiro e econômico e o cenário de crise sanitária mundial imposto pelo vírus da Covid-19, pode-se inferir que houve impacto direto no mercado de odontologia dentro do período citado.

REFERÊNCIAS

ALMEIDA, Renato Neves de. **O método dos mínimos quadrados: estudo e aplicações para o ensino médio**. Campos do Goytacazes: Universidade Estadual do Norte Fluminense Darcy Ribeiro – UENF, 2015.

BOULICAUT, Jean-François; ESPOSITO, Floriana; GIANNOTTI, Fosca; PEDRESCHI, Dino. **Knowledge Discovery in Databases: PKDD**. 8th Conference on Principles and Practice of Knowledge Discovery in Databases. Itália: 2004.

CAMBRONERO, Cristina García; MORENO, Irene Gómez. **Algoritmos de aprendizaje: knn & kmeans**. Intelgencia en Redes de Comunicación, Universidad Carlos III de Madrid, 2006.

CHEIN, Flávia. **Introdução aos modelos de regressão linear**. Brasília: ENAP, 2019.

CLARK, Dan. **Beginning Power BI: A Practical Guide to Self-Service Data Analytics with Excel 2016 and Power BI Desktop**. Apress, 2017.

COELHO, Antônio C.; CUNHA, Jacqueline VA. **Regressão linear múltipla. Análise multivariada: para os cursos de administração, ciências contábeis e economia**. São Paulo: Atlas, 2007.

CUNHA, João Paulo Zanola. **Um estudo comparativo das técnicas de validação cruzada aplicadas a modelos mistos**. Instituto de Matemática e Estatística. São Paulo, 2019.

CUTLER, Adele; CUTLER, D. Richard; STEVENS, John R. **Random forests. Ensemble machine learning: Methods and applications**, 2012.

DEMSAR, Janez; ZUPAN, Blaz. **Orange: Data Mining Fruitful and Fun – A Historical Perspective**. Slovenia: University of Ljubljana, 2013.

DIDATICA. **O que é e como funciona o algoritmo Random Forest**. Disponível em <https://didatica.tech/o-que-e-e-como-funciona-o-algoritmo-randomforest/> . Acesso em 25 out. de 2023.

FACURE, Matheus. **Machine learning tutorials**. Disponível em <https://matheusfacure.github.io>. Acessado em: 24 out. 2023.

FERRARI, Alberto; RUSSO, Marco. **Analyzing Data with Microsoft Power BI and Power Pivot for Excel**. Sample Files on the Web. Washington: Microsoft press, 2017.

FERRERO, Carlos Andres. **Algoritmo kNN para previsão de dados temporais: funções de previsão e critérios de seleção de vizinhos próximos aplicados a variáveis ambientais em limnologia**. 2009. Tese de Doutorado. Universidade de São Paulo.

GOLDBERG, Yoav. **A Primer on Neural Network Models for Natural Language Processing**. Israel: Computer Science Department Bar-Ilan University, 2016.

GUO, Gongde; WANG, Hui; BELL, David; BI, Yaxin; GREER, Kieran. **KNN Model-Based Approach in Classification**. Irlanda do Norte: School of Computing and Mathematics, University of Ulster, 2003.

HASHTAG. **Hashtag Treinamentos**. Disponível em: <https://www.hashtagtreinamentos.com>. Acesso em: 31 out. 2023.

INTUITIVE, Tutorial. **WP content**. Disponível em: <https://intuitivetutorial.com/wp-content>. Acesso em: 24 out. 2023.

KOVÁCS, Zsolt László. **Redes neurais artificiais**. São Paulo: Editora Livraria da Física, 2006.

MACEDO, Sibeli Maria de. **Manual de gestão empresarial para micro e pequenas empresas (MPEs) comerciais: caderno de ferramentas básicas e a construção de dashboards com o Software Power BI**. 2022.

MEDIUM, Portal. **Funcionamento da técnica de Validação cruzada.** Disponível em https://miro.medium.com/v2/resize:fit:1400/1*R3d1vovbUG63DFkeo-NFSw.jpeg . Acesso em: 24 out. 2023.

ORANGE. **Orange Visual Programming.** Disponível em <https://orange3.readthedocs.io>. Acesso em: 27 out. de 2023.

RAMOS, Isabel; SANTOS, Maribel Yasmina. **Business Intelligence: tecnologias da informação na gestão de conhecimento.** Lisboa : FCA Editora de Informática, 2006.

RODRIGUES, Vitor. Métricas de Avaliação: acurácia, precisão, recall... quais as diferenças?. 2019.

SANTOS, Hellen Geremias dos. **Machine learning para análises preditivas em saúde: exemplo de aplicação para predizer óbito em idosos de São Paulo, Brasil.** São Paulo: Cadernos de Saúde Pública, 2019.

SILVEIRA, Guilherme; BULLOCK, Bennett. **Machine Learning: introdução à classificação.** Brasil: Casa do código, 2017.

ZHANG, Cha; MA, Yunqian. **Ensemble Machine Learning: Methods and Applications.** Nova Iorque: Springer, 2012.

AVENI, A. **Sistemas de Saúde e Economia de Saúde – Impactos causados pela covid-19.** Brasil: Cadernos de prospecção, 2020.

SAÚDE, Lean. **O que é a curva ROC, Sensibilidade e Especificidade?** Disponível em <https://www.leansaude.com.br/o-que-e-curva-roc-sensibilidade-e-especificidade/>. Acesso em 12 dez. de 2023.