

PREDIÇÃO DA SÍNDROME RESPIRATÓRIA
AGUDA GRAVE POR MEIO DE
MULTICLASSIFICAÇÃO COM ALGORITMOS DE
MACHINE LEARNING

Data: 05/02/2024

Dedico este trabalho a Deus, minha família,
amigos e colegas que estiveram sempre jun-
tos comigo durante toda minha jornada
acadêmica.

AGRADECIMENTOS

Gostaria de expressar meu sincero agradecimento a todas as pessoas que contribuíram para a conclusão bem-sucedida deste trabalho de conclusão de curso. Neste momento especial, desejo expressar minha gratidão a Deus, minha família, amigos, colegas e meu orientador, além dos professores que estiveram ao meu lado durante essa jornada. Em primeiro lugar, agradeço a Deus por me conceder a força, a sabedoria e a perseverança necessárias para superar os desafios ao longo deste processo. Sua orientação e apoio constantes foram fundamentais para alcançar esse marco importante em minha vida acadêmica.

À minha família, meu porto seguro, expresso minha profunda gratidão. O amor, o incentivo e o apoio incondicionais que recebi de meus pais, irmãos e outros familiares foram cruciais para me manter motivado e focado em meus objetivos. Vocês sempre acreditaram em mim e me encorajaram a nunca desistir, mesmo nos momentos mais difíceis. Sou imensamente grato por ter vocês ao meu lado. Aos meus amigos e colegas, sou grato por compartilharmos essa jornada acadêmica juntos. Agradeço por todas as trocas de conhecimento, apoio mútuo, estudo em grupo e momentos de descontração que compartilhamos ao longo dos anos. Vocês tornaram essa jornada mais leve e memorável. Sou grato pela amizade e pelo suporte inestimável que me proporcionaram.

Ao meu orientador, expresso minha sincera gratidão pela sua orientação, conhecimento e dedicação ao longo do processo de desenvolvimento deste trabalho. Suas orientações valiosas, sugestões e críticas construtivas foram fundamentais para moldar este projeto. Agradeço por ter compartilhado seu tempo, conhecimento e experiência comigo. Sou imensamente grato pela sua paciência e pela confiança que depositou em meu trabalho.

Também quero estender meu agradecimento a todos os professores que contribuíram para a minha formação acadêmica. Suas aulas, orientações e ensinamentos foram fundamentais para meu crescimento intelectual e profissional. Sou grato por ter tido a oportunidade de aprender com profissionais tão dedicados e experientes.

Por fim, gostaria de expressar minha gratidão a todos aqueles que, de alguma forma, contribuíram para o meu sucesso nesta jornada. Seja através de palavras de encorajamento, apoio emocional ou assistência prática, cada gesto foi importante e significativo. Este trabalho não teria sido possível sem a presença de todas essas pessoas especiais em minha vida. Agradeço do fundo do coração por todo o apoio, incentivo e confiança que recebi ao longo dessa jornada. Sinto-me verdadeiramente abençoado por ter cada um de vocês ao meu lado. O meu mais sincero obrigado a Deus, à minha família, amigos, colegas, orientador e professores. Vocês foram essenciais para a minha conquista acadêmica e sou eternamente grato por tudo o que fizeram por mim.

“O senhor é meu pastor e nada me faltará!”
(Salmo 23)

RESUMO

A Síndrome Respiratória Aguda Grave (SRAG) é uma condição médica grave que afeta o sistema respiratório, causando sintomas como febre alta, tosse, falta de ar e dificuldade respiratória. Sua etiologia pode ser diversa, incluindo o coronavírus SARS-CoV-2, responsável pela COVID-19. A SRAG pode levar a complicações graves, como pneumonia e insuficiência respiratória, demandando cuidados hospitalares intensivos e, em casos mais graves, pode resultar em insuficiência de múltiplos órgãos e morte. Este trabalho teve como objetivo demonstrar algumas das principais técnicas de *Machine Learning* para a classificação multiclasse de SRAG, visando aprimorar a capacidade de previsão da síndrome. Para isso, foram utilizados diferentes modelos de classificação, voltados especificamente para classificação multiclasse. Os resultados obtidos revelaram que o modelo XGBoost se destacou, alcançando uma performance geral de 83%. Esse modelo apresentou excelentes resultados na classificação de SRAG por influenza, SRAG por Covid e outras SRAGs relacionadas a vírus respiratórios. No entanto, identificou-se um desempenho razoável nas classes 3 e 4, que representam SRAG por outros agentes etiológicos e um tipo não especificado de SRAG, respectivamente. Esses resultados reforçam a importância do uso de técnicas de *Machine Learning* na análise da SRAG, contribuindo para melhorar a capacidade de previsão e diagnóstico da síndrome. Futuras pesquisas podem se concentrar em aprimorar o desempenho do modelo nas classes mais desafiadoras, buscando aperfeiçoar a aplicação prática dessas técnicas na saúde pública e na medicina.

Palavras-chave: previsão de SRAG; saúde pública; modelos de aprendizado de máquina; classificação multiclasse.

ABSTRACT

Acute Severe Respiratory Syndrome (ASRS) is a severe medical condition that affects the respiratory system, causing symptoms such as high fever, cough, shortness of breath, and respiratory distress. Its etiology can be diverse, including the SARS-CoV-2 coronavirus, responsible for COVID-19. ASRS can lead to serious complications such as pneumonia and respiratory failure, requiring intensive hospital care and, in severe cases, can result in multiple organ failure and death. This study aimed to demonstrate some of the main machine learning techniques for multiclass classification of ASRS, aiming to improve the syndrome's predictive capability. Different classification models specifically tailored for multiclass classification were used to achieve this goal. The results obtained revealed that the XGBoost model stood out, achieving an overall performance of 83%. This model showed excellent results in classifying ASRS caused by influenza, ASRS caused by Covid, and other ASRS related to respiratory viruses. However, reasonable performance was identified in classes 3 and 4, representing ASRS caused by other etiological agents and an unspecified type of ASRS, respectively. These results reinforce the importance of using machine learning techniques in ASRS analysis, contributing to improving the predictive capability and diagnosis of the syndrome. Future research can focus on enhancing the model's performance in the more challenging classes, seeking to refine the practical application of these techniques in public health and medicine.

Keywords: ASRS prediction; public health; machine learning models; multiclass classification.

LIST OF FIGURES

Figura 1 – Funcionamento da Árvore de Decisão	16
Figura 2 – Otimização do Hiperplano	18

LIST OF TABLES

Tabela 1 – Descrição das Variáveis do SRAG Open Data SUS	14
Tabela 2 – Descrição da Proporção dos Dados	19

TABLE OF CONTENTS

1	INTRODUÇÃO	12
2	MATERIAL E MÉTODOS	13
2.1	Material	13
2.1.1	<i>MÉTODOS</i>	15
2.2	Métodos	15
2.2.1	<i>Hiperparâmetros</i>	15
2.3	Árvore de Decisão	16
2.3.1	<i>Critério de Decisão</i>	17
2.3.1.1	<i>Índice de Gini</i>	17
2.4	Suport Vector Machine	17
3	RESULTADOS	19
4	CONCLUSÃO	20
5	REFÊNCIAS	21

1 INTRODUÇÃO

A Síndrome Respiratória Aguda Grave (SRAG), é condição clínica caracterizada por sintomas respiratórios graves que exigem hospitalização e pode ser causada por uma variedade de agentes infecciosos, como vírus, bactérias ou fungos. Ela geralmente se manifesta como uma infecção do trato respiratório inferior, afetando os pulmões e causando sintomas como febre alta, tosse persistente, falta de ar, dor no peito e dificuldade respiratória. Em casos mais graves, a SRAG pode levar à insuficiência respiratória e exigir suporte ventilatório. De acordo com ??) são sintomas do SARGS: indivíduo com síndrome gripal (SG) que apresente: dispneia/desconforto respiratório ou pressão ou dor persistente no tórax ou saturação de oxigênio (O₂) menor que 95% em ar ambiente ou coloração azulada (cianose) dos lábios ou rosto. Vale salientar, no entanto, que nem todos os casos da síndrome são explicados pela infecção por Sars-CoV-2, pois, como mencionado, outras infecções respiratórias podem desencadear o problema (??).

Segundo ??), nos últimos anos, o surgimento da Covid-19 e o aumento no número de casos de gripe geraram uma atmosfera de medo e incertezas. Essa situação é agravada pela semelhança entre os sintomas de diferentes síndromes respiratórias graves, tornando a classificação e o diagnóstico um desafio para os profissionais de saúde. Muitas dessas síndromes compartilham sintomas comuns, como febre, tosse, falta de ar e dor no peito, resultando em uma sobreposição de características clínicas que pode dificultar a identificação precisa da doença em questão.

A importância do *Machine Learning* vem crescendo nos últimos anos, impulsionada pelo aumento na disponibilidade de dados, pela evolução dos algoritmos de aprendizagem e pelo avanço da capacidade de processamento dos computadores. Com isso, tornou-se possível realizar análises cada vez mais sofisticadas e precisas em áreas como finanças, marketing, saúde, entre outras. O uso de técnicas de *Machine Learning* proporciona uma abordagem inovadora para lidar com a complexidade dos dados relacionados à SRAG, permitindo que padrões sejam identificados. Essas técnicas explorarão os dados existentes, como informações clínicas, resultados de exames laboratoriais e históricos médicos, a fim de construir um modelo preditivo capaz de auxiliar na identificação correta do tipo de SRAG.

Nesta monografia, o objetivo é desenvolver um modelo de classificação eficaz utilizando técnicas de aprendizado de máquina *Machine Learning* para identificar e classificar corretamente os diferentes tipos de Síndrome Respiratória Aguda Grave (SRAG). Para alcançar esse objetivo, será empregado o uso de algoritmos de *Machine Learning* que analisarão dados clínicos e epidemiológicos relevantes. O intuito é criar um modelo preditivo capaz de distinguir entre os tipos específicos de SRAG de forma mais precisa e eficiente.

2 MATERIAL E MÉTODOS

Será apresentada uma descrição detalhada dos materiais utilizados, bem como os procedimentos adotados para a coleta e preparação dos dados. Além disso, serão explicadas as principais técnicas de *Machine Learning* empregadas nas análises do problema apresentado nesta monografia, voltadas a classificação multiclasse, abrangendo algumas de suas aplicações específicas nesse contexto desafiador.

2.1 Material

Para a análise, foi utilizado um banco de dados de acesso livre e gratuito do OpenDataSUS (SARG¹) está relacionada aos atributos de pacientes com algum tipo de síndrome gripal e refere-se ao período de 2021 a 2023. O conjunto de dados é formado de 468.994 observações e 166 variáveis: sexo, idade, ID do hospital, entre outras. O dicionário completo pode ser acessado no site oficial do OpenDataSUS ou no anexo desta monografia que contém as descrições das variáveis utilizadas. Na Tabela~2 pode-se observar a descrição de algumas variáveis utilizadas no modelo, incluindo a variável resposta:

¹ <https://opendatasus.saude.gov.br/sq/dataset/srag-2021-a-2023>

Table 1 – Descrição das Variáveis do SRAG Open Data SUS

Variável	Tipo	Categoria	Descrição
AN_PARA3	Categórica	1 - marcado pelo usuário Vazio - não marcado	Resultado do Teste Antigênico. Parainfluenza 3.
AN_ADENO	Categórica	1 - marcado pelo usuário Vazio - não marcado	Resultado do Teste Antigênico. Adenovírus.
AN_OUTRO	Categórica	1 - marcado pelo usuário Vazio - não marcado	Resultado do Teste Antigênico. Outro vírus respiratório.
CLASSI_FIN	Categórica	1 - SRAG por influenza 2 - SRAG por outro vírus respiratório 3 - SRAG por outro agente etiológico, qual: 4 - SRAG não especificado 5 - SRAG por covid-19 pelo usuário	Diagnóstico final do caso.

Fonte: Elaborado pelo Autor, 2023

A variável resposta representa a classificação final do caso, como sendo:

- 1-SRAG por influenza;
- 2-SRAG por outro vírus respiratório;
- 3-SRAG por outro agente etiológico, qual;
- 4-SRAG não especificado;
- 5-SRAG por covid-19.

2.1.1 METODOS

O ambiente computacional estatístico utilizado para a análise e limpeza dos dados foi o *software R* (??) e para a geração dos modelos foi utilizado a *linguagem python* (??), os scripts e os demais arquivos utilizados nas análises desse projeto pode ser acessado através do seguinte repositório do github: [Prediction_of_SARGS_by_multiclassification](#).

2.2 Métodos

Inicialmente foi feita a limpeza dos dados, visando garantir a qualidade e confiabilidade do conjunto de dados utilizado. Em seguida, é realizada uma análise exploratória detalhada, permitindo uma melhor compreensão do nosso conjunto de dados e a identificação de possíveis padrões relevantes. A partir disso, foram utilizadas várias técnicas de *Machine Learning* e Estatística, tais como: métodos de reamostragem e subamostragem, seleção de atributos e alguns modelos de *Machine Learning* para classificação multiclasse. Por último, foi realizada a comparação de desempenho entre os modelos.

No método de subamostragem, foi utilizado a amostragem aleatória, onde instâncias(amostras) aleatórias são selecionadas da classe majoritária para corresponder ao número de amostras da classe minoritária. Isso reduz o tamanho da classe majoritária, equilibrando as proporções das classes. Por último, foram criados vários modelos de *Machine Learning* para classificação multiclasse, usando as técnicas antes descritas, a fim de comparar seus resultados e verificar qual modelo foi capaz de classificar as classes corretamente com maior exatidão e que será o possível candidato a implementação. Além disso, foram feitas a Curva Roc para avaliação do modelo com melhores métricas e o gráfico de importância de atributos Shapley que mostrará o quanto cada covariável é importante para o modelo.

2.2.1 Hiperparâmetros

Segundo BARBOSA (2018), alguns algoritmos de classificação possuem parâmetros que precisam ser configurados antes de executá-los para obtenção de um resultado mais satisfatórios. Os hiperparâmetros são parâmetros ajustáveis externamente que influenciam o comportamento e o desempenho de um modelo de aprendizado de máquina, mas não são aprendidos a partir dos dados durante o processo de treinamento. Ao contrário dos parâmetros do modelo, que são ajustados internamente durante o treinamento, os hiperparâmetros são definidos antes do treinamento e afetam como o modelo é treinado e como ele realiza as previsões. Os hiperparâmetros determinam as configurações e as car-

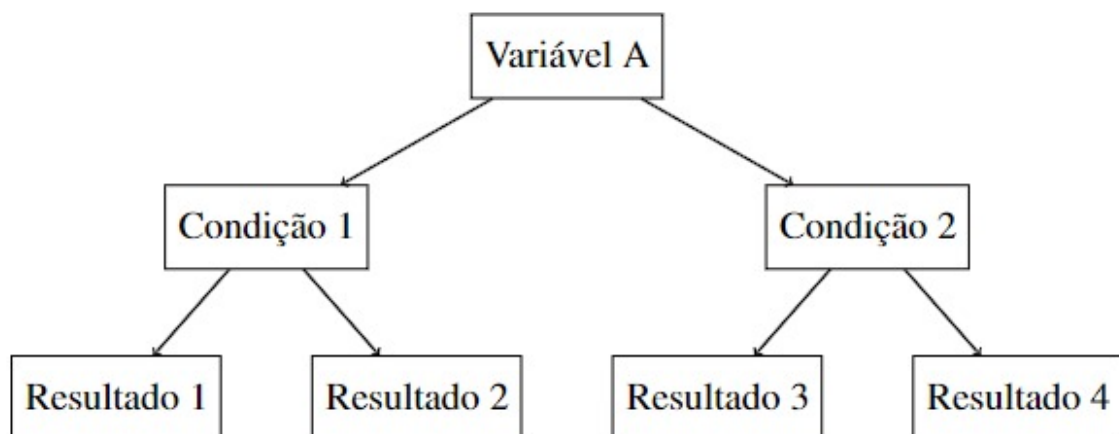
acterísticas do modelo, permitindo que você personalize e ajuste o desempenho conforme as necessidades específicas do problema e do conjunto de dados. Esses hiperparâmetros podem variar conforme o algoritmo e o tipo de modelo utilizado.

2.3 Árvore de Decisão

Segundo Lauretto (2010), uma árvore é uma coleção de elementos chamados nós, dentre os quais um é distinguido como uma raiz, juntamente com uma relação de “paternidade” que impõe uma estrutura hierárquica sobre os nós. Trata-se de um modelo que nos ajuda a tomar decisões com base em uma série de condições e critérios definidos. Sua representação gráfica, que se assemelha a uma árvore com nós e ramos, torna o processo de tomada de decisão mais visual e compreensível.

Imagine que estamos diante de um problema complexo que exige uma série de escolhas para chegar a um resultado desejado. A árvore de decisão nos ajuda a dividir esse problema em pequenas partes e nos guia por diferentes caminhos, com base em critérios específicos. Cada nó da árvore representa uma decisão que precisamos tomar, enquanto os ramos representam as possíveis alternativas que podemos seguir.

Figure 1 – Funcionamento da Árvore de Decisão.



Fonte: Elaborado pelo Autor, 2023

Neste exemplo, é apresentada uma estrutura básica de uma árvore de decisão. O nó raiz representa uma variável de interesse, enquanto os nós subsequentes representam diferentes condições ou critérios de decisão. Cada nó de decisão é conectado a dois ou mais nós de resultado, que representam as possíveis ações ou resultados decorrentes

daquela condição específica. Nas árvores de decisão temos uma métrica chamada critério de decisão.

2.3.1 Critério de Decisão

O critério em árvores de decisão é uma medida utilizada para determinar como os nós da árvore devem ser divididos durante a construção do modelo. O critério define a função de avaliação que determina qual divisão é considerada a melhor em termos de pureza das classes resultantes.

Existem diferentes critérios utilizados em árvores de decisão, sendo os mais comuns o índice de Gini (Gini impurity) e a entropia. Esses critérios medem a impureza dos dados em um nó da árvore e procuram dividir o nó para reduzir a impureza e aumentar a homogeneidade das classes nós filhos.

2.3.1.1 Índice de Gini

O critério de Gini avalia a probabilidade de classificar incorretamente uma amostra aleatória com base na distribuição das classes em um nó. Ele mede a impureza dos dados, sendo que um valor de 0 indica que todas as amostras pertencem a uma única classe e um valor de 1 indica que as amostras estão igualmente distribuídas entre as classes. O índice de Gini pode ser obtido pela seguinte expressão:

$$\text{Gini}(p) = 1 - \sum_{i=1}^K p_i^2,$$

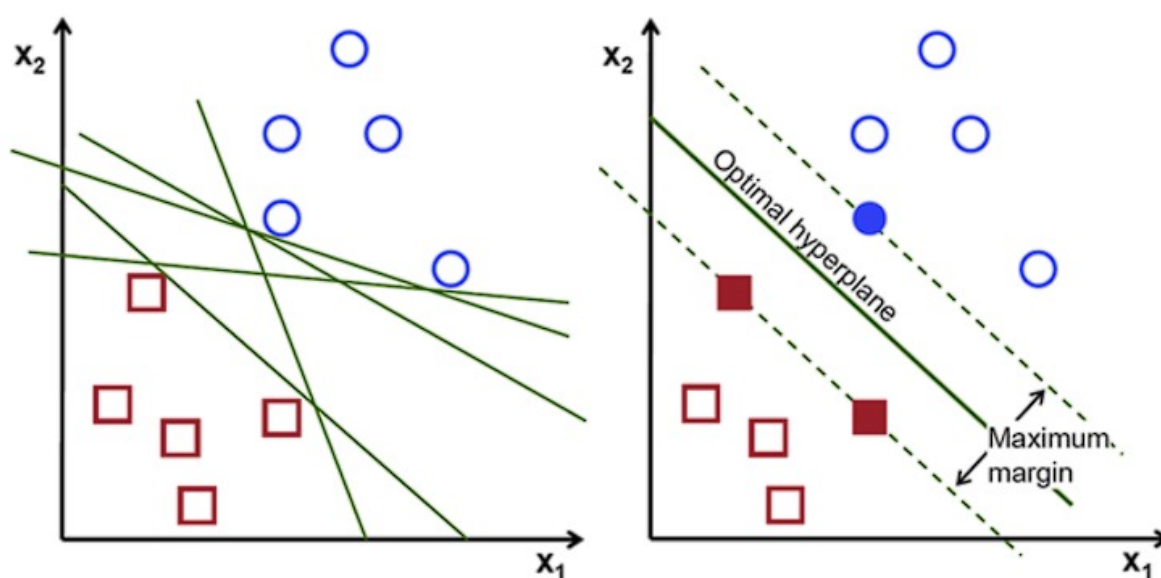
onde, p representa a distribuição de probabilidade das classes em um nó, e K é o número de classes. Para realizar o cálculo do índice de Gini, é necessário seguir alguns passos. Primeiramente, é preciso calcular o quadrado das probabilidades de cada classe, representadas por p_i^2 . Em seguida, esses valores são somados. Por fim, o resultado dessa soma é subtraído de 1.

2.4 Suport Vector Machine

O objetivo do algoritmo da máquina de vetores de suporte (SVM –Support Vector Machine) é encontrar um hiperplano em um espaço N-dimensional (N -o número de recursos ou atributos) que classifica distintamente os pontos de dados, DataScienceAcademy (2021). Segundo Srinivas (2010), é uma técnica de classificação que procura encontrar um

hiperplano que particione os dados por seus rótulos de classe e ao mesmo tempo, evite o ajuste excessivo dos dados maximizando a margem da separação hiperplano. Na Figura 2 podemos visualizar a definição do hiperplano:

Figure 2 – Otimização do Hiperplano



Fonte: Data Science Academy(2022)

3 RESULTADOS

Após a divisão dos dados em dados de espera e dados para treino e teste, no intuito de aumentar as informações sobre a classe minoritária (classe 3), então dividimos mais 50% dos dados de espera, filtramos os valores que são equivalentes a classe 3, uniremos o dataset(banco de dados) que ficou como base reduzida ao dataset filtrando com base na classe 3, após esse procedimento ficamos com a seguinte proporção dos dados:

Table 2 – Descrição da Proporção dos Dados

Classe	Proporção
1	2,33%
2	5,61%
3	1,43%
4	44,92%
5	45,72%

Fonte: Elaborado pelo Autor, 2023

4 CONCLUSÃO

Neste trabalho foram demonstradas algumas das principais técnicas de machine learning para classificação multiclasse, visando aprimorar a capacidade de previsão de síndromes respiratórias agudas graves (SRAG). Ao longo da pesquisa, foram utilizados diferentes modelos de classificação, sendo o XGBoost o que apresentou melhores resultados em relação às métricas de classificação para as cinco classes analisadas. Essa abordagem permitiu uma performance geral de 83%, destacando-se especialmente na classificação de SRAG por influenza, SRAG por Covid, além das SRAGs por outros vírus respiratórios. No entanto, as classes 3 e 4 demonstraram apenas uma performance razoável, ficando ligeiramente acima do nível aleatório. Essa ocorrência pode ser atribuída a diversos fatores. Embora tenhamos realizado o balanceamento dos dados através do undersampling, equilibrando todas as classes, é possível que as classes 1, 2 e 5 tenham definições mais concretas de seus atributos. Em contraste, a classe 3 abrange a SRAG por outros agentes etiológicos, o que pode significar a existência de múltiplos agentes, impactando a distribuição dos atributos e dificultando sua classificação. De maneira similar, a classe 4 refere-se a um tipo de SRAG que não pôde ser especificado com base em seus atributos. Consequentemente, embora o modelo XGBoost tenha apresentado desempenho geral superior em relação aos demais modelos, é necessário considerar o aprimoramento das métricas nas classes 3 e 4. Para prosseguir com a implementação, é recomendável explorar outras técnicas de processamento de dados, bem como estratégias adicionais de balanceamento e ajustes no modelo. Melhorar o desempenho do modelo é crucial, especialmente considerando o contexto da área de saúde.

Após analisar a matriz de confusão do melhor modelo, notou-se que a maioria das predições incorretas das classes 3 e 4 estava concentrada em sua proximidade, indicando uma forte semelhança entre os atributos (características) dessas duas classes. Isso pode ter contribuído para as dificuldades em distingui-las corretamente e resultou em um número maior de classificações equivocadas entre elas. Com tudo, pode-se notar que, mesmo com as ressalvas, o modelo é capaz de prever com excelente performance a classificação de SRAG por influenza, SRAG por Covid, além das SRAGs por outros vírus respiratórios. Vale ressaltar que, quando estamos trabalhando em áreas como a saúde, é importante ter o menor erro possível, pois o erro pode custar vidas.

5 REFERÊNCIAS

- BARBOSA, FELIPE REIS MACEDO. 2018. “Otimização de Hiperparâmetros Em Algoritmos de Árvore de Decisão Utilizando Computação Evolutiva,” 16.
- Bezerra, Francisco Antonio, Jefferson Fernando Grande, and Adriano José da Silva. 2009. “Análise e Caracterização de Modelos de Custos Que Utilizam o Valor de Shapley Para Alocação de Custos Entre Departamentos.” *Gestão & Produção* 16 (1): 74–84. <<https://doi.org/10.1590/S0104-530X2009000100008>>.
- Breslow, N., and J. Crowley. 1974. “A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship.” *The Annals of Statistics*, 437–53.
- Buzaid, Antonio Carlos, Fernando Cotait Maluf, and Debora de Melo Gagliato. 2020. *Vencer o Câncer de Mama*. 2nd ed. São Paulo: Dentrix.
- Câncer - INCA, Instituto Nacional de. 2019. *Câncer de Mama: Vamos Falar Sobre Isso?* 5th ed. Rio de Janeiro.
- . 2022. *Câncer de Mama: Vamos Falar Sobre Isso?* 7th ed. Rio de Janeiro.
- Ceccon, Denny. 2019a. “XGBoost: A Evolução Das Árvores de Decisão.” *IA EXPERT ACADEMY*. <<https://doi.org/2019/04/18/xgboost-a-evolucao-das-arvores-de-decisao/#:~:text=Os%20modelos%20baseados%20em%20%C3%A1rvores,prefer%C3%A2ncia%20dos%20profissionais%20da%20%C3%A1rea.>>
- . 2019b. “XGBoost: A Evolução Das Árvores de Decisão.” <<https://doi.org/2019/04/18/xgboost-a-evolucao-das-arvores-de-decisao/#:~:text=Os%20modelos%20baseados%20em%20%C3%A1rvores,prefer%C3%A2ncia%20dos%20profissionais%20da%20%C3%A1rea.>>
- Colosimo, Enrico Antônio, and Suely Ruiz Giolo. 2006. *Análise de Sobrevida Aplicada*. 1st ed. São Paulo: Blucher.
- Cox, D. R. 1972. “Regression Models and Life-Tables.” *Journal of the Royal Statistical Society. Series B (Methodological)* 34: 187–220.
- . 1975. “Partial Likelihood.” *Biometrika* 62 (2): 269–76.
- Cox, D. R., and E. J. Snell. 1968. “A General Definition of Residuals.” *Journal of the Royal Statistical Society. Series B (Methodological)* 30 (2): 248–75.
- Data, Science Academy. 2022. “Curso: Machine Learning.”
- DataScienceAcademy. 2021. “Curso: machine learning.” <<https://www.datascienceacademy.com.br/course/machine-learning-engineer>>.
- Dias, Jenika Ferreira, Natália Silva Martins, and Clícia Valim Côrtes Gradim. 2018. “Análise de Sobrevida de Mulheres Com Câncer de Mama.” *Revista de Enfermagem UFPE On Line*.
- Espinosa, Raquel, Fernando Jiménez, and José Palma. 2023. *Multi-Surrogate Assisted Multi-Objective Evolutionary Algorithms for Feature Selection in Regression and Classification Problems with Time Series Data*. *Information Sciences*. Vol. 662.

- Fernandez, Alberto, Ronaldo C. Prati, Mikel Galar, Bartosz Krawczyk, Francisco Herrera, and Salvador Garcia. 2018. *Learning from Imbalanced Data Sets. Aprendendo Com Conjuntos de Dados Desequilibrados*. Springer.
- Grambsch, Patricia M., and Terry M. Therneau. 1994. “Proportional Hazards Tests and Diagnostics Based on Weighted Residuals.” *Biometrika* 81 (3): 515–26.
- IARC, INTERNATIONAL AGENCY FOR RESEARCH ON CANCER -. 2020. *Cancer Today*. Lyon: França.
- Jones, Emily. 2018. “Deep Learning Applications in Image Recognition.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (2): 321–35.
- Kaplan, E. L., and Paul Meier. 1958. “Nonparametric Estimation from Incomplete Observations.” *Journal of the American Statistical Association* 53: 457–81.
- Koziarski, Michael. 2020. “Subamostragem Baseada Em Radial Para Classificação de Dados Desequilibrados.” *Reconhecimento de Padrões* 102 (3).
- Lauretto, Marcelo S. 2010. “Árvores de Decisão,” 02.
- Lawless, Jerald F. 1982. *Statistical Models and Methods for Lifetime Data*. 2nd ed. New York: John Wiley; Sons.
- Macedo, Mendes De, and Charles. 2017. “O Que é Machine Learning?”
- Mantel, N. 1966. “Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration.” *Cancer Chemotherapy Reports* 50: 163–70.
- Martins, Vera Lucia Milani, and Liane Werner. 2010. “Análise Não Paramétrica de Falhas Ao Longo Do Calendário Para Alto-Falantes.” *Produto & Produção* 11 (3).
- Morais, Rafael Lima de. 2017. “Uso de Árvores Aleatórias Para Classificação Sensorial de Arroz Cozido.” *Universidade de Brasília*, 22–25.
- Olanda Ferraz, Rosemeire de, and Djalma de Carvalho Moreira Filho. 2017. “Análise de Sobrevivência de Mulheres Com Câncer de Mama: Modelos de Riscos Competitivos.” *Ciência & Saúde Coletiva [Online]* 21 (11).
- Oliveira, Bruna Bianchi. 2021. “Estudo Dos Métodos de Classificação Supervisionada Na Identificação de Malignidade de Tumores Mamários,” 14.
- Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. “Scikit-Learn: Machine Learning in Python.” *Journal of Machine Learning Research* 12: 2825–30.
- Pereira, Paulo Joé, and Mário Javier Ferrua Vivanco. 2003. “Viabilidade Da Aplicação de Mecanismos de Censura Tipo i e Aleatória Em Dados Entomológicos.” *Ciência e Agrotecnologia* 27 (2).
- Python Software Foundation. 2023. *Python Language Reference*. Python Software Foundation. <<https://www.python.org/doc/>>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <<https://www.R-project.org/>>.
- Rosa, Ayrton Lima da. 2019. “Classificação de Imagens de Frutas Utilizando Aprendizado

- de Máquina,” 31.
- Santos, Gustavo Carvalho. 2020. “Algoritmos de Machine Learning Para Previsão de Ações Da B3,” 50.
- Santos, Vanessa Sardinha dos. 2023. “Síndrome Respiratória Aguda Grave (Sars).” <https://brasilecola.uol.com.br/doencas/sindrome-respiratoria-aguda-grave-sars.html>.
- Saúde Brasil, Ministério da. 2021. “Saiba Como é Feita a Definição de Casos Suspeitos de Covid-19 No Brasi.” <https://brasilecola.uol.com.br/doencas/sindrome-respiratoria-aguda-grave-sars.html>.
- Schoenfeld, David. 1982. “Partial Residuals for the Proportional Hazards Regression Model.” *Biometrika* 69: 239–41.
- Shapley, Lloyd S. 1953. “A Value for n-Person Games.” In *Contributions to the Theory of Games*, 2:307–17. 28. Princeton University Press.
- Silva Braga, Ana Cristina da. 2000. “CURVAS ROC: ASPECTOS FUNCIONAIS e APLICAÇÕES,” 32.
- SILVA, RENATA ALVES DA. 2022. “AUTOMEDICAÇÃO RELACIONADA à SÍNDROME GRIPAL ENTRE ADULTOS JOVENS EM TEMPOS DE PANDEMIA DA COVID – 19,” 14.
- Srinivas, Ranjini. 2010. “Managing Large Data Sets Using Support Vector Machines.” <https://Digitalcommons.unl.edu>.
- Strapasson, Elizabeth. 2007. “Comparação de Modelos Com Censura Intervalar Em Análise de Sobrevida.” Doutorado em Agronomia, Universidade de São Paulo - Escola Superior de Agricultura “Luiz de Queiroz.”
- Team, R Core. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org>.
- Therneau, Terry M, Thomas Lumley, Atkinson Elizabeth, and Crowson Cynthia. 2022. *Survival: Survival Analysis*. <https://cran.r-project.org/web/packages/survival>.
- Thuler, Luiz Claudio. 2003. “Considerações Sobre a Prevenção Do Câncer de Mama Feminino.” *Revista Brasileira de Cancerologia* 49 (4): 227–38.