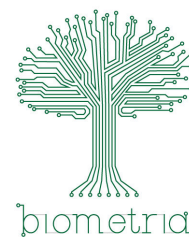


RESSALVA

Atendendo solicitação do(a)
autor(a), o texto completo desta tese
será disponibilizado somente a partir
de 17/08/2024.



Universidade Estadual Paulista “Júlio de Mesquita Filho”
Instituto de Biociências – Câmpus de Botucatu
Programa de Pós-graduação em Biometria



Modelos de predição para dados censurados aplicados a pacientes com doença renal crônica em terapia de substituição renal

Agda Jéssica de Freitas Galletti

Botucatu
2022

Agda Jéssica de Freitas Galletti

Modelos de predição para dados censurados aplicados a pacientes com doença renal crônica em terapia de substituição renal

Tese de Doutorado apresentada ao Curso do Programa de Pós-graduação em Biometria da Universidade Estadual Paulista “Júlio de Mesquita Filho” como parte dos requisitos necessários para a obtenção do título de Doutora em Biometria.

Orientadora: Profa. Dra. Liciana Vaz de Arruda Silveira

Coorientadora: Profa. Dra. Daniela Ponce

Botucatu
2022

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Galletti, Agda Jéssica Freitas.

Modelos de predição para dados censurados aplicados a
pacientes com doença renal crônica em terapia de substituição
renal / Agda Jéssica Freitas Galletti. - Botucatu, 2022

Tese (doutorado) - Universidade Estadual Paulista "Júlio de
Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Liciane Vaz de Arruda Silveira

Coorientador: Daniela Ponce

Capes: 90194000

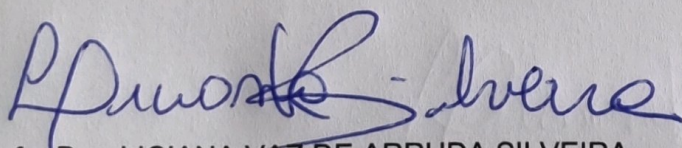
1. Análise de sobrevivência (Biometria). 2. Aprendizado do
computador. 3. Diálise peritoneal. 4. Hemodiálise. 5. Predição
(Logica).

Palavras-chave: Análise de sobrevivência; Aprendizado de
máquina; Diálise peritoneal; Hemodiálise; Modelos de predição.

ATA DA DEFESA PÚBLICA DA TESE DE DOUTORADO DE AGDA JÉSSICA DE FREITAS GALLETTI, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA, DO INSTITUTO DE BIOCIÊNCIAS - CÂMPUS DE BOTUCATU.

Aos 17 dias do mês de agosto do ano de 2022, às 08:00 horas, por meio de Videoconferência, realizou-se a defesa de TESE DE DOUTORADO de AGDA JÉSSICA DE FREITAS GALLETTI, intitulada **Modelos de predição para dados censurados aplicados a pacientes com doença renal crônica em terapia de substituição renal**. A Comissão Examinadora foi constituída pelos seguintes membros: Profa. Dra. LICIANA VAZ DE ARRUDA SILVEIRA (Orientador(a) - Participação Virtual) do(a) Departamento de Bioestatística, Biologia Vegetal, Parasitologia e Zoologia / Instituto de Biociências de Botucatu - UNESP, Prof. Dr. THIAGO SANTOS MOTA (Participação Virtual) do(a) FATEC / Faculdade de Tecnologia de Botucatu, Prof. Dr. MAICON VINICIUS GALDINO (Participação Virtual) do(a) SEDUC / Secretaria da Educação, Profa. Dra. SILVIA HELENA MODENESE GORLA DA SILVA (Participação Virtual) do(a) Coordenadoria de Engenharia Agrônômica / UNESP/Registro, Prof. Dr. ROGERIO ANTONIO DE OLIVEIRA (Participação Virtual) do(a) Departamento de Bioestatística, Biologia Vegetal, Parasitologia e Zoologia / Instituto de Biociências de Botucatu - UNESP. Após a exposição pela doutoranda e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, a discente recebeu o conceito final: _____

APROVADA. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.



Profa. Dra. LICIANA VAZ DE ARRUDA SILVEIRA

À minha amada mãe e aos meus amigos e professores dedico este trabalho. A presença de vocês durante esta incrível e louca jornada tornou tudo mais fácil. Sinto-me grata!
Dedico também a mim por perseverar e persistir ao longo de anos longe de casa na bela e charmosa cidade de Botucatu.

Agradecimentos

Agradeço a Deus e a todos que de alguma forma contribuíram para a realização deste trabalho, em especial:

À minha querida orientadora Profa. Dra. Liciania Vaz por todo suporte e orientação. Além disso, pela amizade, parceria, incansáveis incentivos e conselhos, pelas ótimas conversas e pelas oportunidades de conhecimento a mim ofertadas.

À minha coorientadora Profa. Dra. Daniela Ponce por ter topado embarcar nesta jornada, por todo suporte, parceria e conhecimento compartilhado.

À Unidade de Diálise do Hospital das Clínicas da Faculdade de Medicina de Botucatu e seus pacientes que permitiram o acesso aos prontuários médicos e por todo suporte oferecido.

A todos professores e funcionários do Departamento de Bioestatística, Biologia Vegetal, Parasitologia e Zoologia, Unesp, Botucatu, pelo apoio, gentileza e disponibilidade.

Ao programa de Pós-graduação em Biometria e à Unesp pelas diversas oportunidades a mim oferecidas em docência, por meio de bolsas de estágio à docência e participação dos Programa de Atividades e Aperfeiçoamento em Docência no Ensino Superior (PAADES) e Programa Formação Didático-Pedagógica para Cursos de Modalidade a Distância, que permitiram me tornar uma profissional melhor e ainda contribuíram com a continuidade dos meus estudos.

À minha família que sempre esteve ao meu lado, principalmente à minha mãe por todo apoio, amor e confiança.

A todos os alunos de mestrado e doutorado do programa de pós graduação em Biometria, que direta ou indiretamente, contribuíram positivamente para que este trabalho fosse realizado.

Aos meus amigos que fizeram desta jornada mais agradável e feliz. Obrigada pelos abraços, sorrisos, eventuais momentos de lazer, sessões de cinema, dias e noites de estudos, alimentações saudáveis, besteiras também, festinhas, companheirismo, confiança e inúmeros incentivos. Destacando as minhas albercats e companheiras de casa por serem luz em meio ao caos que foi esta pandemia.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES pelo incentivo a pós-graduações de todos país e por oportunizar a formação de pesquisadores e popularização do conhecimento científico.

Resumo

A doença renal crônica, caracterizada pela alteração da função renal de modo progressivo e irreversível, é não transmissível e uma das principais causas de mortalidade em todo o mundo. Pacientes em estágio muito avançado da doença necessitam iniciar a terapia renal substitutiva, isto é, diálise peritoneal (DP), hemodiálise (HD) ou transplante renal. Segundo a Sociedade Brasileira de Nefrologia, em 2020, 144.779 pacientes estavam em diálise, sendo que, anualmente, cerca de 20 mil novos pacientes iniciaram o tratamento de hemodiálise e apresentam taxa de mortalidade de 15% ao ano. A análise estatística pode auxiliar no planejamento e no desenvolvimento de estratégias para o controle da doença, importante para a manutenção, gestão, prevenção e avaliação de políticas voltadas à saúde. Além disso, a predição do prognóstico de doenças é um dos principais objetivos para médicos e gestores de saúde pública, que pode ser utilizada para direcionamento de intervenções preventivas e podem fornecer a probabilidade do paciente ter ou desenvolver uma determinada doença. Portanto, este trabalho tem como objetivo obter os fatores associados à mortalidade de pacientes em tratamento dialítico por meio de algoritmos de aprendizado de máquina com inclusão dos tempos de sobrevivência censurados. Como resultado, o modelo de Cox teve melhor desempenho preditivo. E também, pacientes que iniciaram a terapia renal substitutiva em hemodiálise têm maior risco de morte do que aqueles que iniciaram com diálise peritoneal. Entretanto, pacientes internados e que iniciaram o tratamento em HD terão menor risco de morte do que aqueles que iniciaram com DP. Além disso, independente do tratamento iniciado, pacientes que foram internados têm maior risco de morte. O modelo de Cox ajustado indicou a idade como um fator de risco, conforme indicado na literatura. Diferentemente, a presença de doenças de base não foram significativas para explicar o tempo de vida dos pacientes em tratamento.

Palavras-chave: aprendizado de máquina; diálise peritoneal; hemodiálise; modelo de predição de sobrevivência.

Abstract

Chronic kidney disease, characterized by progressive and irreversible loss of kidney function, is noncommunicable and one of the main causes of mortality worldwide. Patients in an advanced stage of the disease need to start renal replacement therapy, that is, peritoneal dialysis (PD), hemodialysis (HD) or kidney transplant. According to the Brazilian Society of Nephrology, in 2020, 144,779 patients were on dialysis, with around 20,000 new patients starting hemodialysis annually and a mortality rate of 15% per year. Statistical analysis can assist in the planning and development of strategies for disease control, which is important for the maintenance, management, prevention and evaluation of health policies. In addition, the prediction of disease prognosis is one of the main goals for physicians and public health managers, which can be used to target preventive interventions and can provide the probability of the patient having or developing a certain disease. Therefore, this work aims to obtain the factors associated with mortality of patients on dialysis through machine learning algorithms including censored survival times. As a result, the Cox model had the best predictive performance. Also, patients who started renal replacement therapy on hemodialysis have a higher risk of death than those who started on peritoneal dialysis. However, a patient who is hospitalized and started on HD will have a lower risk of death than a patient who started on PD. In addition, regardless of the treatment initiated, patients who are hospitalized have a higher risk of death. The adjusted Cox model indicated age as a risk factor, according to some references in the literature. Unlike, the presence of underlying diseases was non-significant to explain the lifetime of patients under treatment.

Keywords: machine learning; peritoneal dialysis; hemodialysis; survival prediction model.

Lista de figuras

Figura 1 – Comparação da hemodiálise com a diálise peritoneal	5
Figura 2 – Ilustração de uma árvore de classificação/regressão	18
Figura 3 – Diagrama de rede neural com uma rede oculta	22
Figura 4 – Exemplo de aplicação do MVS destacando os vetores de suporte, a margem e os erros de classificação	26
Figura 5 – Fluxograma de um processo de aprendizado de máquina	28
Figura 6 – Estimativa de Kaplan-Meier da função de sobrevivência	38
Figura 7 – Curvas de sobrevivência estimada via Kaplan-Meier das covariáveis qualitativas	41
Figura 8 – Curvas de sobrevivência estimada via Kaplan-Meier das covariáveis qualitativas	42
Figura 9 – Curvas de sobrevivência estimada via Kaplan-Meier das covariáveis qualitativas segundo terapia de substituição renal	43
Figura 10 – Curvas de sobrevivência estimada via Kaplan-Meier das covariáveis qualitativas segundo terapia de substituição renal	44
Figura 11 – Curvas de sobrevivência estimada via Kaplan-Meier das covariáveis qualitativas segundo terapia de substituição renal	45
Figura 12 – Curvas de sobrevivência estimada via Kaplan-Meier das covariáveis qualitativas segundo terapia de substituição renal	46
Figura 13 – Resíduos padronizados de Schoenfeld do modelo de Cox ajustado	48
Figura 14 – Box plot dos índices C das reamostragens - validação cruzada 10 <i>folds</i> - para os diferentes modelos de sobrevivência	50
Figura 15 – Box plot dos índices C das reamostragens - <i>Bootstrap</i> - para os diferentes modelos de sobrevivência	50
Figura 16 – Previsão da estimativa da sobrevivência segundo cenário	51

Lista de tabelas

Tabela 1	– Estadiamento da doença renal crônica proposto pelo KDOQI e atualizado pelo <i>National Collaborating Centre for Chronic Condition</i>	3
Tabela 2	– Tabela de contingência gerada no tempo t_j	10
Tabela 3	– Hiperparâmetros considerados dos algoritmos que serão otimizados	32
Tabela 4	– Descrição das variáveis observadas	37
Tabela 5	– Distribuição dos pacientes segundo a terapia de substituição renal iniciada e variáveis qualitativas observadas e teste de associação. Botucatu, 2014 – 2019	39
Tabela 6	– Medidas descritivas das variáveis quantitativas observadas segundo a terapia de substituição renal iniciada e teste para comparar os grupos de tratamento. Botucatu, 2014 – 2019	40
Tabela 7	– Resultados dos testes <i>logrank</i> (valor de p) para as comparações dos grupos considerando a TSR	40
Tabela 8	– Estimativas obtidas para o modelo de Cox ajustado aos dados de pacientes renais	47
Tabela 9	– Teste da proporcionalidade dos riscos do modelo de Cox ajustado	47
Tabela 10	– Resultados dos algoritmos otimizados e índice C médio dos modelos considerando os valores otimizados e padrões do pacote	49
Tabela 11	– Desempenho da previsão dos modelos de sobrevivência	49
Tabela 12	– Previsão de risco de óbito de pacientes segundo alguns cenários	51

Lista de abreviaturas e siglas

AIC	Critério de Informação de Akaike.
AM	Aprendizado de máquina.
AS	Análise de sobrevivência.
CART	<i>Classification and Regression Tree</i> (Árvores de Classificação e Regressão).
CVC	Cateter venoso central.
DP	Diálise peritoneal.
DCNT	Doenças crônicas não transmissíveis.
DCR	Doenças renal crônica.
DRET	Doença renal crônica estágio terminal.
EKM	Estimador de Kaplan-Meier.
FAV	Fístula arteriovenosa.
HD	Hemodiálise.
MVS	Máquinas de vetores de suporte.
pmp	Por milhão da população.
RNA	Redes neurais artificiais.
RVS	Regressão do vetor de suporte.
SBN	Sociedade Brasileira de Nefrologia.
TFG	Taxa de filtração glomerular.
TRS	Terapia renal substitutiva.
TSR	Terapia de substituição renal.

Lista de Códigos em R

4.1	Definição do objeto <i>task</i>	29
4.2	Definição do objeto <i>learner</i>	29
4.3	<i>Script</i> - treino, previsão e avaliação do desempenho	29
4.4	Avaliação do desempenho via reamostragem (<i>resampling</i>)	30
4.5	Comparação em modelos - definição do <i>benchmarking</i>	30
4.6	Exemplo - dados <i>Lung</i>	31
4.7	Otimização de hiperparâmetros	34

Sumário

1	INTRODUÇÃO	1
2	DOENÇA RENAL CRÔNICA	3
3	ANÁLISE DE SOBREVIVÊNCIA	7
3.1	Teste <i>logrank</i>	9
3.2	Modelo de Cox	11
3.2.1	Avaliação da proporcionalidade dos riscos	12
4	MODELOS DE PREDIÇÃO	14
4.1	Árvores de Classificação e Regressão	18
4.1.1	Árvores de Sobrevida	20
4.2	Redes Neurais Artificiais	21
4.2.1	Redes Neurais de Sobrevida	24
4.3	Máquinas de Vetores de Suporte	24
4.3.1	Máquinas de Vetores de Suporte de Sobrevida	26
4.4	Aprendizado de Máquina no <i>software</i> R	28
4.4.1	<i>Task</i>	28
4.4.2	Learner	29
4.4.3	Treino, previsão e avaliação do desempenho	29
4.4.4	Reamostragem	30
4.4.5	Comparação de desempenho de modelos	30
4.4.6	Exemplo - dados <i>Lung</i>	31
4.5	Otimização de Hiperparâmetros	32
4.5.1	Exemplo - código R para otimização de hiperparâmetros	34
5	RESULTADOS	36
5.1	Descrição dos dados	36
5.2	Resultados inferenciais	46
5.3	Resultados preditivos	48
6	DISCUSSÃO E CONSIDERAÇÕES FINAIS	52
	REFERÊNCIAS	54

Anexos	61
ANEXO A – PARECER DA COMISSÃO DE ÉTICA EM PESQUISA EM SERES HUMANOS	62
ANEXO B – SAÍDA (<i>OUTPUT</i>) - EXEMPLO - DADOS LUNG	67

1 INTRODUÇÃO

As doenças crônicas não transmissíveis (DCNT) são as principais causas de mortalidade e incapacitação em todo o mundo (SILVA et al., 2017). O Instituto Brasileiro de Geografia e Estatística – IBGE (2014) destaca que as DCNT constituem o problema de saúde de maior magnitude e correspondem por mais de 70% das causas de mortes no Brasil. Segundo o secretário de Vigilância em Saúde do Ministério da Saúde (SVS/MS), Arnaldo Medeiros, durante a Semana das Doenças Crônicas não Transmissíveis, destacou que as DCNT matam cerca de 41 milhões de pessoas a cada ano, isto é, cerca de 71% de todas as mortes no mundo. Sendo que, 77% dessas mortes ocorrem em países de baixa e média renda (BRASIL, 2021b).

Dentre as DCNT, tem-se a doença renal crônica (DRC), caracterizada pela diminuição da função renal, que em muitas vezes tem evolução assintomática. O estágio avançado da doença exige que pacientes necessitem de terapia renal substitutiva (TRS) como a diálise contínua (hemodiálise e diálise peritoneal) ou o transplante de rim (MARINHO et al., 2017). De acordo com Marcelo Mazza, presidente da Sociedade Brasileira de Nefrologia (SBN), a DRC pode ser considerada epidêmica, ao passo que, atinge um a cada dez adultos e a taxa de incidência aumenta cada vez mais.

Anualmente, no Brasil, cerca de 20 mil novos pacientes iniciam o tratamento de hemodiálise e taxa de mortalidade de 15% (BRASILIA, 2020). Segundo o Censo Brasileiro de Diálise 2020, financiado pela SBN, “144.779 pacientes estão em diálise e as taxas estimadas de prevalência e incidência de pacientes por milhão da população (pmp) foram 684 e 209, respectivamente. Dentre os pacientes prevalentes, 92,6% estavam em hemodiálise (HD) e 7,4% em diálise peritoneal (DP)” (NERBASS et al., 2022). Diante da alta prevalência da DRC, a qual cursa com elevada mortalidade, é importante implementar medidas com foco no fortalecimento da atenção primária, conforme indicado pelo secretário de Vigilância em Saúde, Arnaldo Medeiros (BRASIL, 2021b). Estudos relacionados ao tema contribuem para o planejamento, gestão e avaliação de políticas voltadas à saúde e assistência social de pacientes com doença renal crônica sendo a análise estatística desses dados ferramenta fundamental.

O tratamento de algumas enfermidades requer o acompanhamento do paciente até a cura ou o óbito, sendo natural que o acompanhamento de um ou outro paciente seja interrompido, causando uma observação parcial da resposta. Por consequência, a análise de sobrevivência é uma metodologia muito utilizada em pesquisas médicas desta natureza (COLLETT, 2003;

KLEINBAUM; KLEIN, 2012). Além disso, a predição do prognóstico de doenças é um dos objetivos principais para médicos e gestores de saúde pública, que pode ser utilizada para direcionamento de intervenções preventivas, estabelecendo um critério sobre a triagem, tratamento em grupos de alto risco, investigação diagnóstica e escolha de terapia, a fim de evitar o óbito dos pacientes (STEYERBERG, 2009; CHEN, 2020).

A principal característica deste tipo de modelo é a criação de um algoritmo preditivo com boa precisão, consequentemente, ele não precisa ser o correto, ou seja, a análise de resíduos ou da qualidade do ajuste são dispensados (HASTIE; TIBSHIRANI; FRIEDMAN, 2017). Metodologias como a aprendizado de máquina (AM), mecanismo para a busca de padrões e a construção de inteligência em uma máquina para poder aprender, o que implica que poderá ser melhor no futuro a partir de sua própria experiência (GOLLAPUDI, 2016), são ferramentas utilizadas para predição. No processo de aprendizado é levado em conta o tipo de variável a ser estudada, isto é, quando a variável de interesse é quantitativa, tem-se problemas de regressão, quando qualitativa, classificação. É possível encontrar diversos trabalhos como de Maccariello et al. (2008), Segal et al. (2020), Kim et al. (2021) e Monaghan et al. (2021) que optaram em obter modelos preditivos para pacientes renais, entretanto, focados em objetivos distintos e centrados em modelos de classificação ou regressão. No entanto, abordagens para dados de sobrevivência, como serão apresentados, ainda são poucas e seus desenvolvimentos recentes (WANG; LI; REDDY, 2019). Além disso, Spooner et al. (2020) observaram que os resultados obtidos via aprendizado de máquina podem fornecer alternativas mais precisas aos métodos tradicionais de análise de sobrevivência, como o modelo de riscos proporcionais de Cox, principalmente, na presença de dados de alta dimensão. Por estes motivos, o desenvolvimento de modelos de predição é uma interessante abordagem no estudo dos tempos de sobrevivência de pacientes com doença renal crônica.

O objetivo geral da pesquisa foi desenvolver um estudo comparativo entre os modelos de aprendizado de máquina para dados censurados a fim de identificar os fatores de risco ou proteção que interferem na sobrevivência da população de pacientes que apresentam doença renal crônica em tratamento dialítico. Além de apresentar de forma didática como utilizar o *software* R para treinar modelos de aprendizado de máquina de sobrevivência.

Referente a estrutura desta tese, a organização está disposta em seis capítulos. O segundo capítulo apresenta-se um resumo sobre a DRC; no terceiro e quarto capítulos, os conceitos básicos da análise de sobrevivência e algoritmos de aprendizado de máquina com a inclusão da censura. Também será dedicado a apresentação de alguns hiperparâmetros dos modelos de predição e o passo a passo da análise utilizando o *software* R; o quinto capítulo, compõem a análise estatística do conjunto de dados e a descrição dos resultados; e por fim, apresenta-se a discussão e as considerações finais, seguido pelas referências bibliográficas e anexos.

REFERÊNCIAS

- BASTOS, M. G.; KIRSZTAJN, G. M. Doença renal crônica: importância do diagnóstico precoce, encaminhamento imediato e abordagem interdisciplinar estruturada para melhora do desfecho em pacientes ainda não submetidos à diálise. *J Bras Nefrol*, v. 33, n. 1, p. 93–108, 2011. 3, 4
- BASTOS, R. M. R. et al. Prevalência da doença renal crônica nos estágios 3, 4 e 5 em adultos. *Revista da Associação Médica Brasileira*, SciELO Brasil, v. 55, p. 40–44, 2009. 52
- BECKER, M. et al. mlr3 book. URL: <https://mlr3book.mlr-org.com>, 2021. 28, 29, 30
- BISCHL, B. et al. Hyperparameter optimization: Foundations, algorithms, best practices and open challenges. *arXiv preprint arXiv:2107.05847*, 2021. 32
- BISCHL, B. et al. mlr: Machine learning in r. *Journal of Machine Learning Research*, v. 17, n. 170, p. 1–5, 2016. Disponível em: <<https://jmlr.org/papers/v17/15-066.html>>. 28
- BISCHL, B. et al. Resampling methods for meta-model validation with recommendations for evolutionary computation. *Evolutionary computation*, MIT Press, v. 20, n. 2, p. 249–275, 2012. 32
- BRASIL. *Diretrizes clínicas para o cuidado ao paciente com doença renal crônica – DRC no Sistema Único de Saúde*. Brasília: Ministério da Saúde - Secretaria de Atenção à Saúde - Departamento de Atenção Especializada e Temática, 2014. 3, 4
- BRASIL. *Doenças Renais Crônicas (DRC)*. [S.l.]: Ministério da Saúde, 2021. <<https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/d/doencas-renais>>. Acesso em 11/2021. 4
- BRASIL. *Saúde apresenta atual cenário das doenças não transmissíveis no Brasil*. Brasília: Ministério da Saúde, 2021. <<https://www.gov.br/saude/pt-br/assuntos/noticias/2021-1/setembro/saude-apresenta-atual-cenario-das-doencas-nao-transmissiveis-no-brasil>>. Acesso em 10/10/21. 1
- BRASILIA. *Doença renal crônica é epidêmica, diz Sociedade Brasileira de Nefrologia*. Senado Federal: Secretaria de Comunicação Social: Agência Senado, 2020. <<https://senado.jusbrasil.com.br/noticias/820456222/doenca-renal-cronica-e-epidemica-diz-sociedade-brasileira-de-nefrologia>>. Acesso em 15/06/22. 1
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001. 14, 15
- BREIMAN, L. et al. *Classification and Regression Tree*. [S.l.]: Chapman and Hall, 1984. 18, 20, 21
- BRESLOW, N. Contribuição à discussão do artigo de dr cox. *Journal of the Royal Statistical Society, Series B*, v. 34, p. 216–7, 1972. 12
- BURGER, S. V. *Introduction to machine learning with R: Rigorous mathematical analysis*. [S.l.]: O'Reilly Media, Inc., 2018. 14

- CHAUDHARY, K.; SANGHA, H.; KHANNA, R. Peritoneal dialysis first: Rationale. *Clin J Am Soc Nephrol*, v. 6, p. 447—456, 2011. [6](#)
- CHEN, L. Overview of clinical prediction models. *Annals of translational medicine*, AME Publications, v. 8, n. 4, 2020. [2](#)
- CIAMPI, A. et al. Stratification by stepwise regression, correspondence analysis and recursive partition: a comparison of three methods of analysis for survival data with covariates. *Computational Statistics & Data Analysis*, v. 4, n. 3, p. 185–204, 1986. [20](#)
- COLLETT, D. *Modelling Survival Data in Medical Research*. 2nd. ed. Florida: Chapman and Hall/CRC, 2003. 391 p. [1](#), [2](#), [7](#), [9](#), [10](#)
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. São Paulo: Blucher, 2006. 370 p. [7](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#)
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, v. 20, n. 3, p. 273–297, 1995. [24](#)
- COX, D. R. Partial likelihood. *Biometrika*, Oxford University Press, v. 62, n. 2, p. 269–276, 1975. [11](#)
- DIAS, D. B. et al. Peritoneal dialysis as an option for unplanned initiation of chronic dialysis. *Hemodialysis International*, v. 20, n. 4, p. 631–633, 2016. [4](#), [5](#), [6](#)
- DIAS, D. B. et al. Urgent-start peritoneal dialysis: The first year of brazilian experience. *Blood Purif*, v. 44, p. 283—287, 2017. [5](#), [6](#)
- ELGELDAWI, E. et al. Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis. In: MDPI. *Informatics*. [S.l.], 2021. v. 8, n. 4, p. 79. [32](#)
- FARAGGI, D.; SIMON, R. A neural network model for survival data. *Statistics in medicine*, Wiley Online Library, v. 14, n. 1, p. 73–82, 1995. [24](#)
- FERREIRA, E. V. Métodos de reamostragem. *Material de apoio a disciplina de Machine Learning para Cientista de Dados, lecionada na LEG/UFPR*, 2018. [16](#), [17](#)
- FOUODO, C. J. et al. Support vector machines for survival analysis with r. *R Journal*, v. 10, n. 1, 2018. [25](#), [26](#), [27](#)
- FRIEDMAN, J. et al. *The elements of statistical learning*. [S.l.]: Springer series in statistics New York, 2001. v. 1. [21](#), [23](#), [24](#), [26](#)
- GALLETTI, A. J. d. F. *Modelos de fração de cura aplicados aos tempos de sobrevivência de pacientes submetidos à ligadura elástica de varizes no esôfago*. Dissertação (Mestrado em Biometria) — Universidade Estadual Paulista (UNESP), Botucatu, SP, 2018. [9](#)
- GEHAN, E. A. A generalized wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, v. 52, n. 1/2, p. 203–223, 1965. [9](#)
- GERRISH, S. *How smart machines think*. [S.l.]: MIT Press, 2018. [21](#)
- GOLLAPUDI, S. *Practical machine learning*. [S.l.]: Packt Publishing Ltd, 2016. [2](#), [14](#), [15](#), [18](#)

- GONZÁLEZ, A. O. et al. Supervivencia en hemodiálisis vs. diálisis peritoneal y por transferencia de técnica. experiencia en ourense 1976-2012. *Revista de la Sociedad Española de Nefrología*, v. 35, n. 6, p. 562—566, 2015. 6, 52
- GORDON, L.; OLSHEN, R. A. Tree-structured survival analysis. *Cancer Treatment Reports*, p. 1065–1068, 1985. 20
- HARRELL, F. E. et al. Evaluating the yield of medical tests. *Jama*, American Medical Association, v. 247, n. 18, p. 2543–2546, 1982. 17
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2ed. ed. New York, USA: Springer, 2017. 745 p. (Series in Statistics). 2
- HAYKIN, S. *Neural networks and learning machines*, 3/E. New Jersey: Pearson Education, 2009. 906 p. 22, 23
- HAZEWINKEL, A.-D.; GELDERBLOM, H.; FIOCCO, M. Prediction models with survival data: a comparison between machine learning and the cox proportional hazards model. *medRxiv*, Cold Spring Harbor Laboratory Press, 2022. 53
- HECHANOVA, L. A. *Diálise*. [S.l.]: Texas Tech University Health Sciences Center, 2020. <https://www.msdmanuals.com/pt-pt/casa/multimedia/image/KID_hemodialysis_peritoneal_pt>. Acesso em 06/2022. 5
- IBGE. *Pesquisa Nacional de Saúde 2013: Percepção do estado de saúde, estilos de vida e doenças crônicas*. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística - IBGE, 2014. 1
- IVARSEN, P.; POVLSEN, J. V. Can peritoneal dialysis be applied for unplanned initiation of chronic dialysis? *Nephrol Dial Transplant*, v. 29, p. 2201—2206, 2014. 5, 6
- IZBICKI, R.; SANTOS, T. M. dos. *Aprendizado de máquina: uma abordagem estatística*. [S.l.]: Rafael Izbicki, 2020. 15, 22, 23, 24, 25
- JAMES, G. et al. *An introduction to statistical learning*. [S.l.]: Springer, 2013. v. 112. 19
- JIN, H.; LU, Y. Cost-saving tree-structured survival analysis for hip fracture of study of osteoporotic fractures data. *Medical Decision Making*, SAGE Publications Sage CA: Los Angeles, CA, v. 31, n. 2, p. 299–307, 2011. 20
- JIN, H. et al. Alternative tree-structured survival analysis based on variance of survival time. *Medical Decision Making*, v. 24, n. 6, p. 670–680, 2004. 20
- JOHNSON, R. J.; FEEHALLY, J.; FLOEGE, J. *Nefrologia clínica: abordagem abrangente*. [S.l.]: Elsevier Brasil, 2016. 4, 5
- KALBFLEISCH, J. D.; PRENTICE, R. L. *The Statistical Analysis of Failure Time Data*. 2. ed. New Jersey: John Wiley & Sons, 2002. 439 p. (Wiley Series in Probability and Statistics). 9
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, v. 53, n. 282, p. 457–481, 1958. 8
- KATZMAN, J. L. et al. DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology*, BioMed Central, v. 18, n. 1, p. 1–12, 2018. 24

- KATZMAN, J. L. et al. DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology*, BioMed Central, v. 18, n. 1, p. 1–12, 2018. 53
- KIM, H. W. et al. Dialysis adequacy predictions using a machine learning method. *Scientific reports*, Nature Publishing Group, v. 11, n. 1, p. 1–7, 2021. 2, 52
- KLEINBAUM, D. G.; KLEIN, M. *Survival Analysis: A Self-Learning Text*. 3. ed. New York: Springer-Verlag, 2012. 700 p. (Statistics for Biology and Health). 1, 2, 7, 11
- KOREVAAR, J. C. et al. Effect of starting with hemodialysis compared with peritoneal dialysis in patients new on dialysis treatment: a randomized controlled trial. *Kidney Int*, v. 64, n. 6, p. 2222–2228, 2003. 52
- KUHN, M.; JOHNSON, K. *Feature engineering and selection: A practical approach for predictive models*. [S.l.]: CRC Press, 2019. 16
- KVAMME, H.; BORGAN, Ø.; SCHEEL, I. Time-to-event prediction with neural networks and cox regression. *arXiv preprint arXiv:1907.00825*, 2019. 53
- LANG, M. et al. mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software*, dec 2019. Disponível em: <<https://joss.theoj.org/papers/10.21105/joss.01903>>. 28
- LAWLESS, J. F. *Statistical models and methods for lifetime data*. [S.l.]: John Wiley & Sons, 2011. 11
- LEBLANC, M.; CROWLEY, J. Relative risk trees for censored survival data. *Biometrics*, JSTOR, p. 411–425, 1992. 20, 21
- LEE, Y.; BANG, H.; KIM, D. J. How to establish clinical prediction models. *Endocrinol Metab*, v. 31, p. 38–44, 2016. 14
- LUCAS, L. d. S. Árvores, florestas e sua função como preditores: Uma aplicação na avaliação do grau de maturidade de empresas. *Rev. Pmkt*, v. 4, n. 1, p. 6–11, 2011. 18
- MACCARIELLO, E. R. et al. Desempenho de seis modelos de predição prognóstica em pacientes críticos que receberam suporte renal extracorpóreo. *Revista Brasileira de Terapia Intensiva*, SciELO Brasil, v. 20, p. 115–123, 2008. 2, 52
- MANTEL, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Rep.*, v. 50, p. 163–170, 1966. 9, 10
- MARINHO, A. W. G. B. et al. Prevalência de doença renal crônica em adultos no brasil: revisão sistemática da literatura. *Cad. Saúde Colet.*, v. 25, n. 3, p. 379–388, 2017. 1, 3, 6
- MAYER, F. P. Métodos de reamostragem. *Material de apoio a disciplina de Estatística Computacional II, lecionada na LEG/UFPR*, 2021. 16, 17
- MENDES, M. L. et al. Peritoneal dialysis as the first dialysis treatment option initially unplanned. *Brazilian Journal of Nephrology*, SciELO Brasil, v. 39, p. 441–446, 2017. 5, 6
- MEYER, D. et al. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. [S.l.], 2021. R package version 1.7-9. Disponível em: <<https://CRAN.R-project.org/package=e1071>>. 33

- MOISEN, G. Classification and regression trees. In: *Jørgensen, Sven Erik; Fath, Brian D. (Editor-in-Chief). Encyclopedia of Ecology, volume 1. Oxford, UK: Elsevier. p. 582-588., p. 582-588, 2008. 19*
- MONAGHAN, C. K. et al. Machine learning for prediction of patients on hemodialysis with an undetected sars-cov-2 infection. *Kidney360*, American Society of Nephrology, v. 2, n. 3, p. 456, 2021. 2, 52
- MORGAN, J. N.; SONQUIST, J. A. Problems in the analysis of survey data, and a proposal. *Journal of the American statistical association*, Taylor & Francis, v. 58, n. 302, p. 415-434, 1963. 18
- MUELLER, J. P.; MASSARON, L. *Machine learning for dummies*. [S.l.]: John Wiley & Sons, 2021. 15, 25
- NERBASS, F. B. et al. Censo brasileiro de diálise 2020. *Brazilian Journal of Nephrology*, SciELO Brasil, 2022. 1
- PEREIRA, E. R. S. et al. Prevalence of chronic renal disease in adults attended by the family health strategy. *Brazilian Journal of Nephrology*, SciELO Brasil, v. 38, p. 22-30, 2016. 52
- PONCE, D.; BRABOA, A. M.; BALBIA, A. L. Urgent start peritoneal dialysis. *Wolters Kluwer Health*, v. 27, 2018. 5, 6
- PROBST, P.; BOULESTEIX, A.-L.; BISCHL, B. Tunability: importance of hyperparameters of machine learning algorithms. *The Journal of Machine Learning Research*, JMLR. org, v. 20, n. 1, p. 1934-1965, 2019. 32
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>. 28
- ROJAS, R. *Neural networks: a systematic introduction*. [S.l.]: Springer Science & Business Media, 2013. 23
- SBN. *Diálise Peritoneal*. [S.l.]: Sociedade Brasileira de Nefrologia, 2021. <<https://www.sbn.org.br/orientacoes-e-tratamentos/tratamentos/dialise-peritoneal/>>. Acesso em 10/2021. 4
- SBN. *Hemodiálise*. [S.l.]: Sociedade Brasileira de Nefrologia, 2021. <<https://www.sbn.org.br/orientacoes-e-tratamentos/tratamentos/hemodialise/>>. Acesso em 10/2021. 4
- SBN. *Compreendendo os rins*. [S.l.]: Sociedade Brasileira de Nefrologia, 2022. <<https://www.sbn.org.br/o-que-e-nefrologia/compreendendo-os-rins/>>. Acesso em 05/2022. 3
- SBN. *Doença renal crônica: diagnóstico e prevenção*. [S.l.]: Sociedade Brasileira de Nefrologia, 2022. <<https://www.sbn.org.br/noticias/single/news/doenca-renal-cronica-diagnostico-e-prevencao/>>. Acesso em 05/2022. 4
- SCHMID, M.; WRIGHT, M. N.; ZIEGLER, A. On the use of harrell's c for clinical risk prediction via random survival forests. *Expert Systems with Applications*, Elsevier, v. 63, p. 450-459, 2016. 17
- SEGAL, Z. et al. Machine learning algorithm for early detection of end-stage renal disease. *BMC nephrology*, Springer, v. 21, n. 1, p. 1-10, 2020. 2, 52

SESSO, R. C. C. et al. Diálise crônica no brasil - relatório do censo brasileiro de diálise. *J Bras Nefrol*, v. 34, n. 3, p. 272–277, 2012. 52

SHIVASWAMY, P. K.; CHU, W.; JANSCHKE, M. A support vector approach to censored targets. In: IEEE. *Seventh IEEE international conference on data mining (ICDM 2007)*. [S.l.], 2007. p. 655–660. 26, 27

SILVA, A. R. et al. Doenças crônicas não transmissíveis e fatores sociodemográficos associados a sintomas de depressão em idosos. *J Bras Psiquiatr.*, v. 66, n. 1, p. 45–51, 2017. 1

SONABEND, R. *survivalmodels: Models for Survival Analysis*. [S.l.], 2022. R package version 0.1.11. Disponível em: <<https://CRAN.R-project.org/package=survivalmodels>>. 33

SONABEND, R. et al. mlr3proba: An r package for machine learning in survival analysis. *Bioinformatics*, 02 2021. ISSN 1367-4803. 28, 29

SPOONER, A. et al. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Scientific reports*, Nature Publishing Group, v. 10, n. 1, p. 1–10, 2020. 2, 52

STEYERBERG, E. W. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York: Springer, 2009. 497 p. 2, 14

TACONELI, C. *Árvores de classificação multivariadas fundamentadas em coeficientes de dissimilaridade e entropia* 100 p. 100 p. Tese (Tese de Doutorado) — Tese (Doutorado em Estatística e Experimentação Agrônômica)—Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP, 2008. 18, 19

TAY, K. *What is Harrell's C-index?* 2019. <<https://statisticaloddsandends.wordpress.com/2019/10/26/what-is-harrells-c-index/>>. Acesso em 06/2022. 17

TERMORSHUIZEN, F. et al. Hemodialysis and peritoneal dialysis: comparison of adjusted mortality rates according to the duration of dialysis: analysis of the netherlands cooperative study on the adequacy of dialysis. *J Am Soc Nephrol.*, v. 14, n. 11, p. 2851–2860, 2003. 52, 53

THERNEAU, T.; ATKINSON, B. *rpart: Recursive Partitioning and Regression Trees*. [S.l.], 2022. R package version 4.1.16. Disponível em: <<https://CRAN.R-project.org/package=rpart>>. 33

THERNEAU, T. M. *A Package for Survival Analysis in R*. [S.l.], 2022. R package version 3.3-1. Disponível em: <<https://CRAN.R-project.org/package=survival>>. 31

THERNEAU, T. M.; GRAMBSCH, P. M. *Modeling Survival Data: Extending the Cox Model*. 1. ed. [S.l.]: Springer, 2000. (Statistics for Biology and Health). ISBN 9781441931610; 1441931619; 9781475732948; 1475732945. 13

THERNEAU, T. M.; GRAMBSCH, P. M.; FLEMING, T. R. Martingale-based residuals for survival models. *Biometrika*, v. 77, n. 1, p. 147–160, 1990. 20

VAPNIK, V. *The nature of statistical learning theory*. [S.l.]: Springer science & business media, 1999. 24, 27

VAPNIK, V.; LERNER, A. Y. Recognition of patterns with help of generalized portraits. *Avtomat. i Telemekh*, v. 24, n. 6, p. 774–780, 1963. 24

VONESH, E. F. et al. Mortality studies comparing peritoneal dialysis and hemodialysis: What do they tell us? *Kidney International*, v. 70, p. S3—S11, 2006. [53](#)

WANG, P.; LI, Y.; REDDY, C. K. Machine learning for survival analysis: A survey. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 51, n. 6, p. 1–36, 2019. [2](#), [20](#)

YILDIRIM, S. Hyperparameter tuning for support vector machines-c and gamma parameters. URL: <https://towardsdatascience.com/hyperparameter-tuning-for-support-vector-machines-c-and-gamma-parameters-6a5097416167>, 2020. [33](#)

Anexos

A Parecer da Comissão de Ética em Pesquisa em Seres Humanos