

ESCOLA DE CIÊNCIAS DA SAÚDE E DA VIDA
CURSO DE PSICOLOGIA

GUILHERME AUGUSTO DIAS

**APRENDIZAGEM DE MÁQUINA PARA PREDIÇÃO DE DESFECHOS EM SAÚDE MENTAL:
UMA INTRODUÇÃO PARA PSICÓLOGOS**

Porto Alegre
2024

GRADUAÇÃO



Pontifícia Universidade Católica
do Rio Grande do Sul

GUILHERME AUGUSTO DIAS

**APRENDIZAGEM DE MÁQUINA PARA PREDIÇÃO DE
DESFECHOS EM SAÚDE MENTAL: UMA INTRODUÇÃO
PARA PSICÓLOGOS**

Trabalho de Conclusão de Curso apresentado
como requisito para a obtenção do grau de Ba-
charel em Psicologia pela Escola de Ciências
da Saúde e da Vida da Pontifícia Universi-
dade Católica do Rio Grande do Sul.

Orientador: Prof. Dr. Wagner de Lara Machado

Porto Alegre

2024

GUILHERME AUGUSTO DIAS

APRENDIZAGEM DE MÁQUINA PARA PREDIÇÃO DE DESFECHOS EM SAÚDE MENTAL: UMA INTRODUÇÃO PARA PSICÓLOGOS

Trabalho de Conclusão de Curso apresentado
como requisito para a obtenção do grau de Ba-
charel em Psicologia pela Escola de Ciências
da Saúde e da Vida da Pontifícia Universi-
dade Católica do Rio Grande do Sul.

Aprovado em 17 de julho de 2024.

Banca examinadora:

Prof. Ma. Caroline Santa Maria Rodrigues – PUCRS

Prof. Dra. Clarissa Pinto Pizarro de Freitas – PUCRS

Prof. Dr. Wagner de Lara Machado – PUCRS

Agradecimentos

Agradeço aos meus pais, Ademilson José Dias e Irinês Porfírio Dias, pelo apoio incondicional que me deram ao longo da vida, incentivando mesmo as ideias questionáveis como aquela de fazer uma segunda graduação.

A minha irmã, Maria Eduarda Dias, cuja trajetória acadêmica é fonte de inspiração.

A minha namorada, Nathalia Fallavena Ceratti, que acompanhou a esta jornada de perto, oferecendo apoio e compreensão que fizeram toda a diferença.

Aos meus amigos, Guilherme da Silva Petró, Norton Neves e Uilinton Luciano Marques da Silva, que foram compreensíveis nos finais de semana em que não pude jogar RPG.

Ao professor Wagner de Lara Machado pela orientação e apoio no desenvolvimento deste trabalho.

Resumo

Desenvolvimentos na área de aprendizagem de máquina representam uma série de oportunidades para diversas áreas de pesquisa e de atuação profissional. O uso de modelos de aprendizagem de máquina na predição de desfecho para tratamentos em saúde mental tem sido investigado e apresenta resultados promissores. A aplicação desse tipo de modelo possibilitaria, por exemplo, maior assertividade na personalização de tratamentos psicoterápicos a nível individual. Apesar dos benefícios potenciais, o tema ainda é pouco discutido nos espaços de pesquisa, formação e atuação em psicologia clínica. Este trabalho tem por objetivo introduzir o tema de aprendizagem de máquina a psicólogos. Busca-se apresentar, de maneira acessível, os conceitos e terminologia básicos em aprendizagem de máquina, proporcionando um panorama geral da área. Para complementar a exposição conceitual, é construído um exemplo a partir de dados de um estudo sobre eficácia de intervenções em psicologia positiva para depressão via internet. Por fim, discutem-se algumas implicações do uso desse tipo de tecnologia para a pesquisa e para a prática em psicoterapia.

Palavras-chave: aprendizagem de máquina, psicoterapia, psicologia positiva, depressão.

Abstract

Advances in machine learning have opened up a multitude of opportunities for various research and professional fields. The use of machine learning models to predict treatment outcomes in mental health has been investigated showing promising results. The application of such models would allow greater assertiveness in tailoring psychotherapeutic treatments at the individual level. Despite its potential benefits, the topic remains unexplored in the fields of research, training and practice in clinical psychology. This work aims to introduce the topic of machine learning to psychologists. It presents the basic machine learning concepts and terminology in an accessible manner, providing a comprehensive overview of the field. To complement the conceptual discussion, an example is constructed using data from a study on the effectiveness of internet-based positive psychology interventions for depression. Finally, implications of using this type of technology in psychotherapy research and practice are discussed.

Keywords: machine learning, psychotherapy, positive psychology, depression.

Lista de ilustrações

Figura 1 – Um exemplo de árvore de decisão construída manualmente.	18
Figura 2 – Modelo de árvore de decisão gerado.	21
Figura 3 – Classificações para o conjunto de dados de teste.	22

Sumário

1	INTRODUÇÃO	9
2	CONCEITOS BÁSICOS DE APRENDIZAGEM DE MÁQUINA . .	11
2.1	O que é aprendizagem de máquina?	11
2.2	Os tipos de aprendizagem de máquina	12
2.2.1	Aprendizagem supervisionada, não supervisionada e por reforço	12
2.2.2	Modelos discriminativos e generativos	13
2.3	A construção de uma aplicação de aprendizagem de máquina	13
2.3.1	Pré-processamento	13
2.3.2	Separação dos dados	14
2.3.3	Treinamento	14
2.3.4	Avaliação do desempenho	15
3	CONSTRUÇÃO DE UM EXEMPLO	16
3.1	O estudo original	16
3.1.1	Participantes	16
3.1.2	Intervenções	16
3.1.3	Controle	17
3.1.4	Desfechos	17
3.2	Árvores de decisão	17
3.3	Plano de análise de dados	19
3.4	Resultados e discussão	20
4	CONCLUSÃO	23
	REFERÊNCIAS	24

1 Introdução

Diversos modelos de psicoterapia apresentam evidência de eficácia no tratamento de transtornos mentais. Uma parcela significativa dos pacientes, no entanto, não responde às intervenções, podendo até mesmo apresentar piora quando submetida a tratamento psicoterápico (Cuijpers et al., 2021). A variabilidade na taxa de resposta ao tratamento é parcialmente explicada pelo quadro clínico do paciente; ainda assim, observou-se que pessoas de uma mesma população clínica respondem de maneiras diferentes a um mesmo tratamento (Hofmann et al., 2012). Especula-se que um atendimento personalizado a nível individual possa melhorar a taxa de resposta ao tratamento (Norcross; Wampold, 2010; Norcross; Lambert, 2018).

A literatura aponta uma série de variáveis preditoras de desfecho clínico para pacientes em tratamento psicoterápico, o que permitiria a adaptação dos protocolos de intervenção de acordo com a expectativa de resposta (Smagula et al., 2019; Andover et al., 2020). Não existe, porém, consenso acerca do valor preditivo de cada variável a nível individual e interações complexas entre diferentes preditores podem prejudicar a acurácia das previsões e impactar propostas de personalização de tratamento negativamente (Taubitz; Budenbender; Alpers, 2022).

Técnicas de inteligência artificial baseadas em aprendizagem de máquina apresentam a capacidade de integrar uma grande quantidade de dados, impondo poucas restrições ao comportamento das variáveis observadas e produzindo modelos flexíveis aplicáveis em diferentes contextos (Dwyer; Falkai; Koutsouleris, 2018). Em essência, a aprendizagem de máquina consiste no uso de métodos estatísticos e computacionais para identificar padrões de relacionamento subjacentes a um grande conjunto de dados, permitindo a construção de modelos discriminativos ou generativos (Roth et al., 2018). Estudos sobre a aplicação de modelos de aprendizagem de máquina na previsão do desfecho de tratamentos em saúde mental apresentam resultados promissores (Dwyer; Falkai; Koutsouleris, 2018), o que possibilitaria maior assertividade na personalização de tratamentos psicoterápicos a nível individual.

Embora representem um potencial ganho para a prática em psicologia clínica, as tecnologias de aprendizagem de máquina em saúde mental não são apresentadas durante a formação de profissionais da psicologia. As diretrizes nacionais curriculares para cursos de graduação em psicologia propõe um modelo formativo generalista, deixando pouco espaço para a apresentação de conhecimentos metodológicos muito específicos como a aprendizagem de máquina (Brasil, 2023; Rudá; Coutinho; Almeida Filho, 2019).

Este trabalho tem por objetivo apresentar os conceitos básicos relacionados à

aprendizagem de máquina, de modo que possa servir como uma introdução acessível ao tema. Busca-se também ilustrar o uso da aprendizagem de máquina no auxílio à tomada de decisões para o planejamento de intervenções psicoterápicas. Utilizando um conjunto de dados de uma intervenção digital em psicologia positiva para depressão (Collins et al., 2023), pretende-se construir um modelo de árvore de decisão para prever o desfecho da intervenção para novos pacientes.

2 Conceitos básicos de aprendizagem de máquina

2.1 O que é aprendizagem de máquina?

Aprendizagem de máquina é a área da ciência da computação que tem como objetivo geral o desenvolvimento de programas de computador capazes de aprender a realizar uma tarefa sem serem explicitamente programados (Bi et al., 2019; Theobald, 2021). Neste contexto, aprendizagem refere-se a aplicação de procedimentos estatísticos e computacionais sobre um conjunto de informações empíricas, buscando alcançar melhorias de desempenho em uma determinada tarefa (Theobald, 2021).

Aprender trata-se, portanto, de ajustar os parâmetros de um modelo estatístico e computacional aos dados observados de modo a maximizar o desempenho na tarefa em questão. Esse processo de aprendizagem é comumente chamado de treinamento do modelo (Bi et al., 2019). Programas de computador baseados em aprendizagem de máquina são capazes de identificar padrões de interação complexos entre variáveis em conjuntos de dados com alta dimensionalidade para realizar tarefas de classificação, regressão, agrupamento e outras (Theobald, 2021).

Considere, por exemplo, um estudo observacional hipotético que investiga a relação entre características de personalidade e o nível de satisfação profissional entre psicólogos. O estudo baseia-se no modelo dos cinco grandes fatores da personalidade (Hutz; Bandeira; Trentini, 2018) e usa o instrumento da Bateria Fatorial da Personalidade para coleta de dados, registrando as pontuações obtidas nas escalas de neuroticismo, extroversão, socialização, realização e abertura (Nunes; Hutz; Nunes, 2015). Além disso, os participantes do estudo reportam o próprio nível de satisfação profissional em uma escala que contém os seguintes valores: baixo, médio e alto.

É possível utilizar esse conjunto de dados para construir um modelo de aprendizagem de máquina preditivo. Um algoritmo processa o conjunto de dados, identificando os padrões de interação existentes entre as variáveis preditoras (características de personalidade) e o desfecho de interesse (nível de satisfação profissional). O conhecimento adquirido durante o processamento dos dados é codificado nos parâmetros de um modelo de aprendizagem de máquina. O modelo pode então ser utilizado para fazer previsões sobre o nível de satisfação profissional de um indivíduo qualquer a partir de suas características de personalidade.

2.2 Os tipos de aprendizagem de máquina

As técnicas de aprendizagem de máquina podem ser organizadas de diferentes maneiras, incluindo classificação pela estratégia adotada durante o processo de aprendizagem e pelo objetivo geral de aprendizagem (Theobald, 2021; Ng; Jordan, 2001).

2.2.1 Aprendizagem supervisionada, não supervisionada e por reforço

As categorias mais comumente usadas na descrição de modelos de aprendizagem de máquina dizem respeito à estratégia de aprendizagem adotada. O modelo pode ser construído segundo uma abordagem de aprendizagem supervisionada, aprendizagem não supervisionada ou aprendizagem por reforço (Theobald, 2021; Bi et al., 2019).

A aprendizagem supervisionada assemelha-se ao processo de aprendizagem adotado por seres humanos, onde o aprendiz identifica padrões a partir de um conjunto de exemplos preparado por um tutor. Durante a fase de aprendizagem, o modelo é exposto a um conjunto de dados que contém informações sobre o desfecho de interesse para cada uma das observações. O acesso às informações de desfecho providas por um agente externo confere o caráter de supervisão a este processo. Técnicas de aprendizagem de máquina para regressão e classificação (support vector machines, árvores de decisão, redes neurais) pertencem a esta categoria (Theobald, 2021; Bi et al., 2019). Um exemplo para a aplicação deste tipo de aprendizagem é usar de dados de ensaios clínicos, onde o desfecho para cada paciente é conhecido, na construção de um modelo capaz de prever o resultado da intervenção para novos pacientes (Collins et al., 2023).

Na aprendizagem não supervisionada, o conjunto de dados analisado não contém qualquer informações sobre desfecho de interesse. Espera-se que o modelo identifique os padrões de relacionamento existentes entre as variáveis do conjunto de dados e gere agrupamentos ou projeções de maneira autônoma. Técnicas de aprendizagem de máquina para tarefas de agrupamento e redução de dimensionalidade (k-means clustering, PCA, TSNE) pertencem a esta categoria (Theobald, 2021; Bi et al., 2019). Um exemplo para a aplicação deste tipo de aprendizagem é investigar os padrões de comorbidade em uma determinada população clínica (Sánchez-Rico; Alvarado, 2019).

Na aprendizagem por reforço, o modelo aprende através de repetidos ciclos de tentativa e erro. A cada ciclo de aprendizagem, o modelo recebe feedback sobre seu desempenho na tarefa, o feedback é incorporado à base de conhecimento construída pelo modelo em ciclos passados e, assim, melhora seu desempenho progressivamente (Theobald, 2021; Bi et al., 2019). Um exemplo para a aplicação deste tipo de aprendizagem é auxiliar a tomada de decisões de tratamento em condições crônicas como a esquizofrenia (Shortreed et al., 2010).

2.2.2 Modelos discriminativos e generativos

Estratégias de aprendizagem supervisionada e não supervisionada podem ser utilizada na construção de modelos com objetivos de aprendizagem distintos. Modelos discriminativos tem por objetivo modelar probabilidade condicional de um desfecho dadas determinadas condições (Bi et al., 2019; Ng; Jordan, 2001). Um modelo discriminativo, poderia representar diretamente a probabilidade de resposta a uma intervenção psicoterápica dadas as condições socioeconômicas do paciente, como escolaridade e renda. Modelos discriminativos são comumente usados em tarefas de regressão e classificação (Bi et al., 2019).

Modelos generativos buscam modelar a distribuição de probabilidade conjunta para as variáveis presentes no conjunto de dados, ou seja, a probabilidade associada a cada combinação de variáveis observada no conjunto de dados de treinamento (Bi et al., 2019; Ng; Jordan, 2001). Um modelo generativo poderia, por exemplo, representar a probabilidade associada a cada combinação de escolaridade, renda e resposta à intervenção observada durante seu treinamento. A distribuição de probabilidade conjunta completa representa, em certa medida, o processo subjacente de geração dos dados, o que permite que modelos generativos sejam utilizados para gerar observações sintéticas (Bi et al., 2019). Esse tipo de modelo é associado a ferramentas de inteligência artificial generativa como o Chat GTP (Wu et al., 2023).

2.3 A construção de uma aplicação de aprendizagem de máquina

O processo para construção de modelos de aprendizagem de máquina pode variar de acordo com a abordagem adotada, mas, de modo geral, consiste na sequência de etapas de pré-processamento, separação dos dados, treinamento e avaliação de desempenho do modelo (Greener et al., 2021).

2.3.1 Pré-processamento

O desempenho de um modelo de aprendizagem de máquina depende, em grande medida, da forma como o conjunto de dados é apresentado. Assim é fundamental uma etapa de processamento inicial para garantir que os padrões mínimos de qualidade de dados são atendidos. Tarefas de pré-processamento comuns são imputação de dados faltantes, balanceamento de classes através de *up-sampling* ou *down-sampling*, recodificação de variáveis categóricas e padronização de variáveis quantitativas (Delgadillo, 2020).

2.3.2 Separação dos dados

Deve-se avaliar o desempenho do modelo resultante ao final do processo de aprendizagem. Uma avaliação efetiva deve verificar o comportamento do modelo quando exposto a um conjunto de dados inéditos, permitindo uma boa estimativa de seu desempenho em um contexto naturalístico. Assim, uma parte dos dados disponíveis, cerca de 10%, deve ser reservada para a avaliação de desempenho. Esses dados são chamados conjunto de dados de teste e não devem ser usados em nenhuma das etapas de treinamento (Greener et al., 2021).

O processo de treinamento do modelo deve ser monitorado para evitar falhas de aprendizagem. Por exemplo, generalizações indevidas feita a partir de uma única observação. O monitoramento é feito a partir de avaliações de desempenho intermediárias, que acontecem durante o treinamento. É necessário, portanto reservar uma parte dos dados restantes, cerca de 10%, para fazer o monitoramento. Esses dados são chamados de dados de validação (Greener et al., 2021).

Uma alternativa para aumentar quantidade de dados disponíveis para treinamento é o uso da técnica de validação cruzada, sendo a *k-fold cross-validation* sua apresentação mais comum. Nesta abordagem, o conjunto de dados de treinamento é dividido em k partes de tamanhos iguais e são realizadas k rodadas de treinamento; a cada rodada uma das partes é separada e usada como conjunto de dados de validação. Assim, evita-se a reserva de dados exclusivos para validação (Delgadillo, 2020).

2.3.3 Treinamento

O treinamento consiste na implementação e execução do algoritmo de aprendizagem de máquina responsável por treinar o modelo. A implementação geralmente faz uso de uma linguagem de programação de alto nível e um framework de aprendizagem de máquina (Greener et al., 2021). Um exemplo de framework é o scikit-learn (Scikit-learn,) para a linguagem de programação Python (Python Software Foundation,).

O processo de treinamento envolve também a seleção de hiperparâmetros: parâmetros de configuração do algoritmo de treinamento. Hiperparâmetros recebem este nome para diferenciá-los dos parâmetros do modelo em construção. Os hiperparâmetros selecionados governam o processo de aprendizagem e podem ter um impacto importante no desempenho do modelo. A seleção de hiperparâmetros pode ser manual ou automatizada durante o processo de treinamento por meio da técnica de *k-fold cross-validation*, onde, em cada uma das rodadas de treinamento, uma configuração de hiperparâmetros é utilizada e avaliada (Delgadillo, 2020).

2.3.4 Avaliação do desempenho

A avaliação do desempenho de um modelo varia de acordo com a estratégia de aprendizagem adotada. Na abordagem de aprendizagem supervisionada é possível expor o modelo a um novo conjunto de dados e comparar os resultados apresentados com os desfechos reais (Greener et al., 2021). Modelos de aprendizagem supervisionada que realizam tarefas de regressão são avaliados com medidas de erro como o erro quadrático médio (Delgadillo, 2020). Modelos de aprendizagem supervisionada para tarefas de classificação utilizam métricas de acurácia preditiva como a área sob a curva ROC (Greener et al., 2021).

Para modelos de aprendizagem não supervisionada, a informação de desfecho não é conhecida; assim é preciso utilizar medidas de ajuste do modelo aos dados de treinamento. A avaliação do desempenho de modelos de aprendizagem não supervisionada é feita comumente com a utilização de medidas de dispersão, homogeneidade e variância preservada (Naeem et al., 2023).

3 Construção de um exemplo

Usando dados de um ensaio clínico, construiu-se um modelo de aprendizagem de máquina para prever o nível de resposta à intervenção sendo testada. Durante o processo de construção, optou-se pela utilização de técnicas simples, colocando a didática acima do desempenho preditivo.

3.1 O estudo original

O estudo original buscava avaliar o impacto de intervenções de psicologia positiva conduzidas via internet sobre a percepção de felicidade e sintomas depressivos (Woodworth et al., 2017), uma tentativa de replicar os resultados obtidos em um trabalho anterior conduzido por Seligman e colaboradores (Seligman et al., 2005).

3.1.1 Participantes

Os participantes foram recrutados por meio de anúncios em veículos de comunicação australianos: páginas web, jornais e uma estação de rádio local. Um total de 295 participantes completou a fase inicial de pré-teste. O grupo era composto majoritariamente por mulheres (85,06%), com idades entre 18 e 83 anos ($M = 43,76$; $SD = 12,43$); a maior parte dos participantes possuía nível de educação superior (74,88%) e classificou a própria renda como média ou acima da média (76%) (Woodworth et al., 2017; Collins et al., 2023).

3.1.2 Intervenções

Os participantes foram distribuídos aleatoriamente em quatro grupos, três grupos experimentais e um grupo de controle; cada um dos três grupos experimentais recebeu uma intervenção distinta. O primeiro grupo experimental recebeu a intervenção de *visita da gratidão*: os participantes foram instruídos escrever e entregar uma carta de agradecimento para alguém que lhes tivesse sido gentil no passado. O segundo grupo experimental recebeu a intervenção de *três coisas boas*: os participantes deveriam anotar três coisas boas que haviam acontecido durante seu dia, justificando suas escolhas de eventos. O último grupo experimental recebeu a intervenção de *pontos fortes*: os participantes receberam uma intervenção psicoeducativa sobre forças de caráter e foram aconselhados a buscar maneiras criativas para utilizar suas próprias forças de caráter no cotidiano. Todas as intervenções tiveram duração de uma semana (Woodworth et al., 2017).

3.1.3 Controle

O grupo de controle foi exposto à atividade placebo de *memórias de infância*: os participantes receberam a instrução de reservar um momento ao final do dia para escrever sobre suas memórias de infância durante uma semana (Woodworth et al., 2017).

3.1.4 Desfechos

O estudo avaliou percepção de felicidade por meio do AHI (Authentic Happiness Inventory), um instrumento de autorrelato composto por 24 itens. Os itens são pontuados em uma escala de que vai de 1 a 5 e a pontuação total é calculada pela soma das pontuações obtidas para cada item. Resultados maiores representam maiores níveis de felicidade (Park; Park; Peterson, 2010).

Sintomas depressivos foram mensurados com a escala CES-D (Center for epidemiologic studies depression scale), um instrumento de autorrelato composto por 20 itens que avaliam a frequência de sintomas em uma escala Likert de 4 pontos. A CES-D é dividida em quatro subescalas: afeto deprimido, afeto positivo, queixas somáticas e problemas interpessoais. A pontuação em uma subescala é calculada pela soma das pontuações obtidas para cada item associado à subescala. A pontuação total é calculada pela soma das pontuações obtidas para cada subescala. Resultados maiores representam maior severidade dos sintomas depressivos (Radloff, 1977).

O estudo também coletou informação sobre sexo, idade, nível de escolaridade e renda das participantes. As medidas foram obtidas nas ocasiões de pré-teste, pós-teste e 4 ocasiões de follow-up: uma semana após o pós-teste, um mês após o pós-teste, três meses após o pós-teste e seis meses após o pós-teste (Woodworth et al., 2017).

3.2 Árvores de decisão

Árvores de decisão representam uma classe de modelos de aprendizagem de máquina que pode ser usada para tarefas de classificação e regressão, ou seja, são capazes de fazer predições para valores de variáveis categóricas e numéricas. Elas utilizam uma série de regras de decisão para prever o desfecho de interesse; essas regras consistem em verificações feitas sobre os valores de variáveis presentes no conjunto de dados usados no treinamento do modelo (Theobald, 2021; Bi et al., 2019).

Na figura 1, tem-se uma árvore de decisão construída manualmente que realiza a tarefa de predição da resposta a uma intervenção psicoterápica hipotética. Com acesso às informações necessárias, é possível navegar pela estrutura da árvore e prever o desfecho que a intervenção teria para uma nova paciente. Para isso, aplicam-se as regras de decisão sucessivamente a partir do nó inicial, chamado raiz, até alcançar um nó terminal, chamado

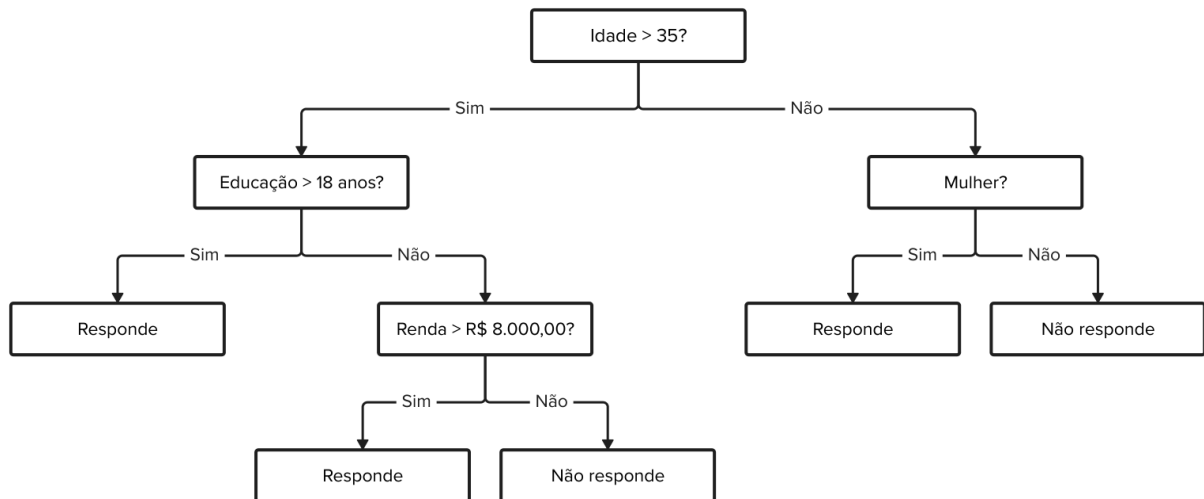


Figura 1 – Um exemplo de árvore de decisão construída manualmente.

folha, que contém a informação de desfecho. Por exemplo, uma paciente de 38 anos, que estudou durante 15 anos de sua vida e tem uma renda mensal de R\$ 10.000,00 teria uma previsão de desfecho com resposta positiva à intervenção.

Algoritmos de aprendizagem de máquina para construção de árvores de decisão operam por meio de um processo iterativo: o processamento do conjunto de dados de treinamento acontece em ciclos que se repetem um determinado número de vezes. Partindo do conjunto de dados de treinamento completo, geram-se as regras de decisão possíveis e seleciona-se aquela que produz partições mais homogêneas dos dados em relação ao desfecho de interesse. As partições geradas servem como base para o próximo ciclo de processamento. O processo segue até que se alcance um critério de parada pré-estabelecido, como o tamanho da árvore ou um número mínimo de observações em nós terminais (Bi et al., 2019).

O uso de árvores de decisão apresenta uma série de benefícios. Árvores de decisão exigem um volume de dados menor para realização do treinamento quando comparadas a outras classes de modelos como as redes neurais artificiais (Theobald, 2021). Elas são capazes de modelar relações não lineares entre as variáveis do conjunto de dados, desempenhando melhor que modelos de regressão linear em contextos onde esse tipo de interação existe (Bi et al., 2019). Além disso, árvores de decisão são facilmente interpretáveis; elas possuem uma representação gráfica que pode ser compreendida de maneira intuitiva, facilitando sua inspeção e a obtenção de insights sobre o processo de tomada de decisão empregado pelo modelo (Bi et al., 2019). As limitações de modelos de árvore de decisão incluem sua sensibilidade a pequenas perturbações no conjunto de dados, como outliers, e uma inclinação ao sobreajuste: uma adaptação excessiva às nuances do conjunto de dados de treinamento que prejudica a capacidade de generalização das previsões feitas pelo modelo (Bi et al., 2019).

3.3 Plano de análise de dados

Construiu-se um modelo de aprendizagem de máquina do tipo árvore de decisão para prever melhora nos níveis de sintomas depressivos a nível individual após a intervenção. Foram utilizados a linguagem de programação Python na versão 3.12 (Python Software Foundation,) e o pacote para processamento estatístico e de aprendizagem de máquina scikit-learn na versão 1.5 (Scikit-learn,).

O conjunto de dados do estudo original, que incluía observações sobre diversas ocasiões de follow-up, foi filtrado para obter somente as observações de pré-teste e pós-teste; os dados de pré-teste e pós-teste de cada participante foram combinados para compor uma única observação. A partir dos dados de pré-teste, foram mantidas as variáveis sociodemográficas (sexo, idade, nível de educação e renda), pontuação em cada item do instrumento AHI, pontuação total no instrumento AHI, pontuação em cada item do instrumento CES-D e pontuação total no instrumento CES-D. Para os dados de pós-teste, foi mantida a variável de pontuação total no instrumento CES-D. Observações relacionadas a participantes do grupo de controle foram descartadas. Após os tratamentos iniciais, o conjunto de dados disponíveis para a construção do modelo tinha 102 observações sendo cerca de 70% de não respondentes.

Adicionou-se uma variável catagórica para indicar melhora nos níveis de sintomas depressivos. A variável teve seu valor preenchido a partir da variação na pontuação total no instrumento CES-D: participantes que apresentaram um aumento de cinco pontos no pós-teste em relação ao pré-teste receberam a classificação de respondentes ao tratamento; os demais participantes receberam a classificação de não respondentes.

Separou-se aleatoriamente 20% das observações disponíveis para compor o conjunto de dados de teste; os 80% restantes foram utilizados para o treinamento do modelo de árvore de decisão. O critério estabelecido para a seleção de regras de decisão foi o de entropia de Shannon (Scikit-learn,) e a profundidade máxima permitida para a árvore resultante foi de cinco níveis.

Utilizou-se o modelo na predição de desfechos para o conjunto de dados de testes, mantido em separado até então, e seu desempenho foi avaliado por meio das métricas de acurácia, precisão e recall. Acurácia representa o percentual de predições corretas realizadas pelo modelo de modo geral. Precisão refere-se à razão entre classificações positivas corretas e o total de classificações positivas feitas pelo modelo. Recall refere-se à razão entre as classificações positivas corretas e o total de observações positivas no conjunto de dados de teste.

3.4 Resultados e discussão

O modelo de árvore de decisão gerado é apresentado na figura 2. O modelo conta com 17 nós organizados em uma estrutura com cinco níveis de profundidade sendo um nó raiz, sete nós intermediários e nove nós folha. As regras de decisão selecionadas verificam os valores de oito variáveis distintas: a pontuação total no CES-D, a pontuação nos itens 13 e 19 do CES-D, a pontuação nos itens 7, 15, 20 e 22 do AHI e a renda da participante. O item 13 da CES-D diz respeito a falar menos que o usual; o item 19 do instrumento refere-se ao sentimento de não ser apreciado por outras pessoas. Os itens 7, 15, 20 e 22 da AHI referem-se respectivamente a sentimentos de tédio, satisfação com o trabalho, bom uso do tempo e experiências de prazer e dor. Além da regra de decisão selecionada, o diagrama apresenta, para cada nó, informações sobre as partições geradas durante o processo de treinamento: o número de observações que alcançaram o nó (*samples*), o nível de entropia de Shannon para as observações (*entropy*), a distribuição das observações entre as classes não respondente e respondente respectivamente (*value*) e a classe predominante nas observações (*class*). A coloração dos nós no diagrama indica a classe predita pelo modelo, com a cor laranja representando participantes não respondentes e a cor azul representando participantes respondentes.

A figura 3 apresenta a matriz de confusão com as classificações obtidas para o conjunto de dados de teste. O modelo construído foi capaz de prever a melhora nos níveis de sintomas depressivos das participantes do conjunto de dados de teste com uma acurácia de 0,714, precisão de 0,500 e recall de 0,833.

Neste exemplo, buscou-se prever melhora nos níveis de sintomas depressivos após uma intervenção de psicologia positiva via internet usando um modelo de aprendizagem de máquina do tipo árvore de decisão. Foi possível, mesmo com um conjunto de dados de tamanho pequeno, construir um modelo funcional e com desempenho razoável, alcançando cerca de 70% de acurácia. O resultado obtido alinha-se com uma série de achados recentes, onde, com técnicas similares, as características sociodemográficas e clínicas de pacientes foram usadas para prever o desfecho de psicoterapia com sucesso (Collins et al., 2023; Hornstein et al., 2021). A incorporação desse tipo de tecnologia na prática de psicoterapia permitiria a personalização de tratamentos com maiores níveis de assertividade: com auxílio de modelos de aprendizagem de máquina, o psicoterapeuta poderia estimar as chances de sucesso de diversas intervenções a um custo relativamente baixo, escolhendo aquela que julgar mais adequada.

O modelo construído possui uma representação gráfica intuitiva que evidencia os critérios usados no processo de classificação das observações. A capacidade de inspecionar o funcionamento do modelo é uma característica altamente desejável, pois permite que o psicoterapeuta use informações adicionais sobre o contexto do paciente para avaliar a qualidade da predição feitas pelo modelo, aumentando a confiança nas decisões clínicas

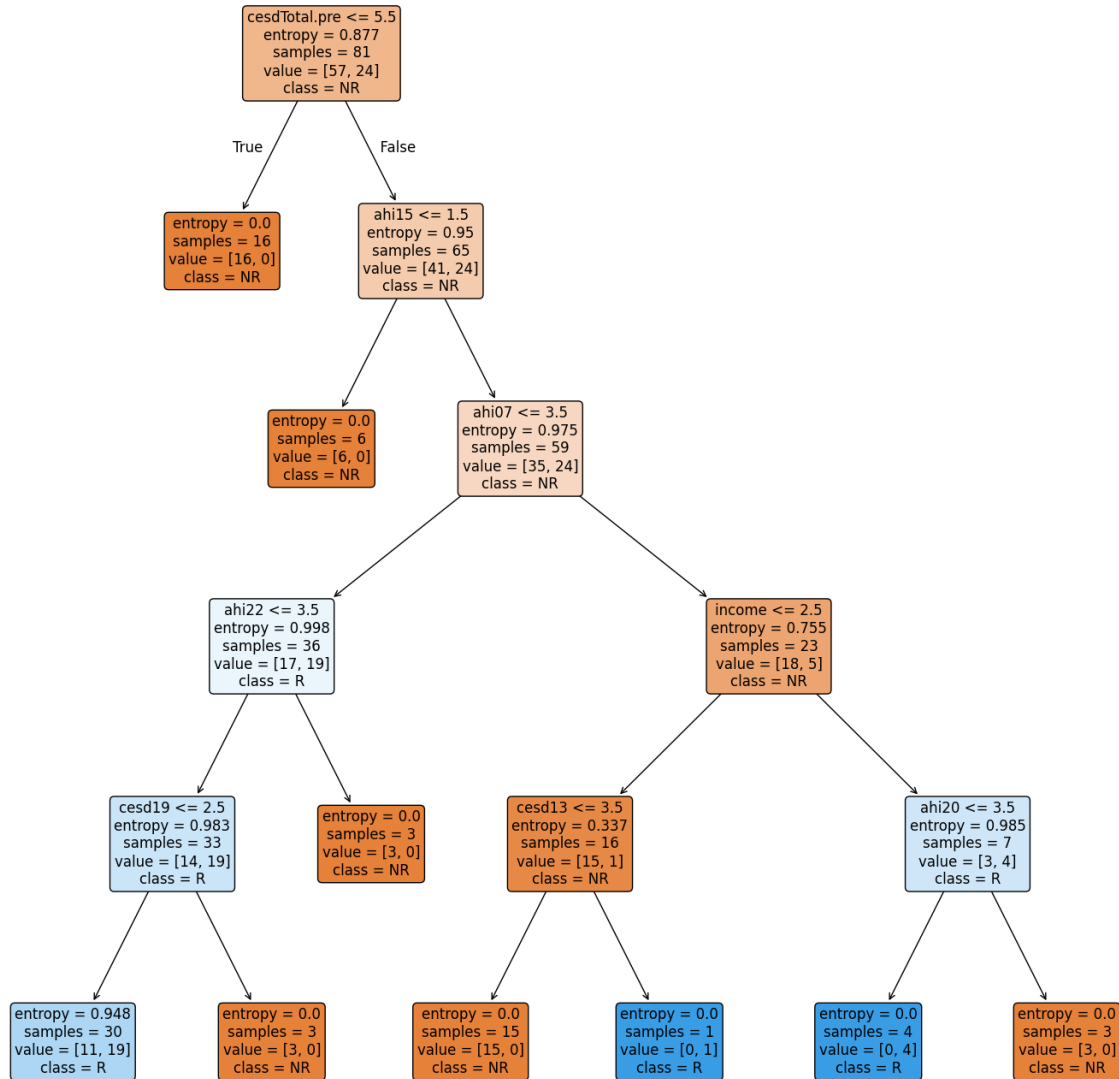


Figura 2 – Modelo de árvore de decisão gerado.

tomadas com auxílio desse tipo de tecnologia (Stiglic et al., 2020; Organization, 2023). A capacidade de interpretar as predições é uma característica particular de alguns tipos de modelos, como as árvores de decisão; outras classes de modelos de aprendizagem de máquina não podem ser facilmente inspecionadas. As redes neurais artificiais, por exemplo, são uma classe de modelos de aprendizagem de máquina que apresenta alta capacidade de aprendizagem e grande poder preditivo, mas não oferecem transparência em relação a seu funcionamento (Stiglic et al., 2020). A possibilidade de inspecionar classes de modelos mais complexos é um problema de pesquisa em aberto e uma condição necessária para a implantação da tecnologia no contexto de atuação profissional em saúde mental (Stiglic et al., 2020; Organization, 2023).

Observa-se que o modelo final apresenta um viés de classificação que favorece

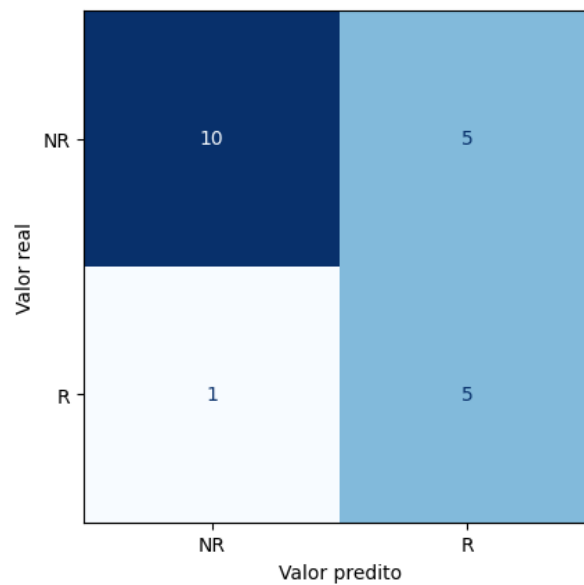


Figura 3 – Classificações para o conjunto de dados de teste.

a produção de falsos positivos, a atribuição da classe respondente a pacientes que não respondem ao tratamento. O viés fica evidente na métrica de precisão de 50%: apenas metade das predições positivas produzidas pelo modelo estavam corretas. O surgimento do viés pode ser atribuído, em parte, ao pequeno número de observações disponíveis para o treinamento do modelo. Além disso, grande parte das observações correspondiam a pacientes não respondentes, o que pode ter prejudicado a capacidade do modelo de identificar pacientes respondentes (Delgadillo, 2020). Embora algumas técnicas de balanceamento de classes possam mitigar o problema (Bi et al., 2019), uma solução definitiva envolve necessariamente a obtenção de um conjunto de dados de treinamento maior e com distribuição equilibrada entre as classes de desfecho (Delgadillo, 2020). O uso de modelos com esse tipo de viés poderia resultar em um grande número de pacientes encaminhados para intervenções com poucas chances de sucesso. O encaminhamento falho tem consequências econômicas e clínicas negativas como o desperdício de recursos financeiros, o adiamento de um tratamento eficaz e a exposição a potenciais efeitos negativos do tratamento (John; Sharma, 2009).

4 Conclusão

Este trabalho teve como objetivo principal introduzir o tema de aprendizagem de máquina para psicólogos. Foi exposto um panorama geral com a apresentação dos conceitos fundamentais da área. Buscando complementar a exposição conceitual, construiu-se um exemplo concreto a partir de dados de um estudo de eficácia de intervenção em psicologia positiva para depressão.

Observou-se que o desenvolvimento de ferramentas de aprendizagem de máquina cada vez mais acessíveis e a ampla disseminação do conhecimento relacionado à área contribuem positivamente para o avanço desta tecnologia no campo da psicologia. O uso de tecnologias baseadas em aprendizagem de máquina no auxílio a tomada de decisões clínicas em psicoterapia mostra-se uma possibilidade tecnicamente viável com potencial de melhorar a qualidade e eficiência dos serviços oferecidos.

Uma das principais limitações do estudo é o acesso a um conjunto de dados de treinamento com volume e qualidade adequados. Técnicas avançadas de pré-processamento e treinamento poderiam, em parte, mitigar o impacto desta limitação, mas seu uso foi descartado com o objetivo de manter a construção simples e didática. Assim, a escassez de dados para treinamento do modelo limitou o desempenho do modelo construído durante a construção do exemplo.

Trabalhos futuros incluem a criação de diretrizes para coleta e avaliação da qualidade de dados para desenvolvimento de modelos de aprendizagem de máquina na psicologia, atuando diretamente sobre um dos maiores desafios da pesquisa na área, a obtenção de dados adequados. Deve-se também investir no desenvolvimento de mecanismos de inspeção para modelos complexos de alto desempenho como as redes neurais artificiais, uma condição necessária para a utilização da tecnologia na prática clínica.

O avanço no desenvolvimento e disseminação das tecnologias de aprendizagem de máquina representa uma oportunidade de melhoria na pesquisa e atuação profissional em psicoterapia. Deve-se investir na adaptação das técnicas ao contexto de tratamentos em saúde mental para realizar o potencial benéfico que a aplicação da tecnologia de aprendizagem de máquina oferece.

Referências

- ANDOVER, M. S. et al. Moderators of treatment response to an intervention for nonsuicidal self-injury in young adults. **Journal of Consulting and Clinical Psychology**, American Psychological Association (APA), v. 88, n. 11, p. 1032–1038, nov. 2020. ISSN 0022-006X.
- BI, Q. et al. What is machine learning? a primer for the epidemiologist. **American Journal of Epidemiology**, Oxford University Press (OUP), Baltimore, out. 2019. ISSN 1476-6256.
- BRASIL. Ministério da Educação. Conselho Nacional de Educação. Câmara de Educação Superior. Resolução CNE/CES nº 1, de 11 de outubro de 2023. **Diário Oficial União: Seção 1**, ed. 201, Brasília, p. 55, 2023.
- COLLINS, A. C. et al. Predicting individual response to a web-based positive psychology intervention: a machine learning approach. **The Journal of Positive Psychology**, Informa UK Limited, v. 19, n. 4, p. 675–685, set. 2023. ISSN 1743-9779.
- CUIJPERS, P. et al. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. **Acta Psychiatrica Scandinavica**, Wiley, v. 144, n. 3, p. 288–299, jul. 2021. ISSN 1600-0447.
- DELGADILLO, J. Machine learning: A primer for psychotherapy researchers. **Psychotherapy Research**, Informa UK Limited, Crestwood, v. 31, n. 1, p. 1–4, dez. 2020. ISSN 1468-4381.
- DWYER, D. B.; FALKAI, P.; KOUTSOULERIS, N. Machine learning approaches for clinical psychology and psychiatry. **Annual Review of Clinical Psychology**, Annual Reviews, v. 14, n. 1, p. 91–118, maio 2018. ISSN 1548-5951.
- GREENER, J. G. et al. A guide to machine learning for biologists. **Nature Reviews Molecular Cell Biology**, Springer Science and Business Media LLC, v. 23, n. 1, p. 40–55, set. 2021. ISSN 1471-0080.
- HOFMANN, S. G. et al. The efficacy of cognitive behavioral therapy: A review of meta-analyses. **Cognitive Therapy and Research**, Springer Science and Business Media LLC, v. 36, n. 5, p. 427–440, jul. 2012. ISSN 1573-2819.
- HORNSTEIN, S. et al. Predicting therapy outcome in a digital mental health intervention for depression and anxiety: A machine learning approach. **DIGITAL HEALTH**, SAGE Publications, v. 7, p. 205520762110606, jan. 2021. ISSN 2055-2076.
- HUTZ, C. S.; BANDEIRA, D. R.; TRENTINI, C. M. **Avaliação Psicológica da Inteligência e da Personalidade**. 1. ed. Porto Alegre: Artmed, 2018.
- JOHN, H.; SHARMA, V. Misdiagnosis of bipolar disorder as borderline personality disorder: clinical and economic consequences. **The World Journal of Biological Psychiatry**, Informa UK Limited, v. 10, n. 4-2, p. 612–615, jan. 2009. ISSN 1814-1412.

- NAEEM, S. et al. An unsupervised machine learning algorithms: Comprehensive review. **International Journal of Computing and Digital Systems**, Deanship of Scientific Research, v. 13, n. 1, p. 911–921, abr. 2023. ISSN 2210-142X.
- NG, A.; JORDAN, M. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In: DIETTERICH, T.; BECKER, S.; GHAHRAMANI, Z. (Ed.). **Advances in Neural Information Processing Systems**. [S.l.]: MIT Press, 2001. v. 14.
- NORCROSS, J. C.; LAMBERT, M. J. Psychotherapy relationships that work iii. **Psychotherapy**, American Psychological Association (APA), v. 55, n. 4, p. 303–315, dez. 2018. ISSN 0033-3204.
- NORCROSS, J. C.; WAMPOLD, B. E. What works for whom: Tailoring psychotherapy to the person. **Journal of Clinical Psychology**, Wiley, v. 67, n. 2, p. 127–132, nov. 2010. ISSN 1097-4679.
- NUNES, C. H. S. d. S.; HUTZ, C. S.; NUNES, M. F. O. **Bateria Fatorial da Personalidade - BFP: manual técnico**. 2. ed. São Paulo: Casa do Psicólogo, 2015.
- ORGANIZATION, W. H. Publications. **Regulatory considerations on artificial intelligence for health**. [S.l.]: World Health Organization, 2023. xiv, 61 p. p.
- PARK, N.; PARK, M.; PETERSON, C. When is the search for meaning related to life satisfaction? **Applied Psychology: Health and Well-Being**, Wiley, v. 2, n. 1, p. 1–13, fev. 2010. ISSN 1758-0854.
- PYTHON SOFTWARE FOUNDATION. Welcome to python.org. Disponível em: <<https://www.python.org/>>. Acesso em: 26-06-2024.
- RADLOFF, L. S. The ces-d scale: A self-report depression scale for research in the general population. **Applied Psychological Measurement**, SAGE Publications, v. 1, n. 3, p. 385–401, jun. 1977. ISSN 1552-3497.
- ROTH, J. A. et al. Introduction to machine learning in digital healthcare epidemiology. **Infection Control and Hospital Epidemiology**, Cambridge University Press (CUP), v. 39, n. 12, p. 1457–1462, nov. 2018. ISSN 1559-6834.
- RUDÁ, C.; COUTINHO, D.; ALMEIDA FILHO, N. d. Formação em psicologia: uma análise curricular de cursos de graduação no brasil. **Revista e-Curriculum**, Informa UK Limited, v. 17, n. 2, p. 419–440, jun. 2019. ISSN 1809-3876.
- SÁNCHEZ-RICO, M.; ALVARADO, J. M. A machine learning approach for studying the comorbidities of complex diagnoses. **Behav. Sci. (Basel)**, MDPI AG, v. 9, n. 12, p. 122, nov. 2019.
- SCIKIT-LEARN. **scikit-learn: machine learning in Python**. Disponível em: <<https://scikit-learn.org/>>. Acesso em: 26-06-2024.
- SELIGMAN, M. E. P. et al. Positive psychology progress: Empirical validation of interventions. **American Psychologist**, American Psychological Association (APA), v. 60, n. 5, p. 410–421, jul. 2005. ISSN 0003-066X.

- SHORTREED, S. M. et al. Informing sequential clinical decision-making through reinforcement learning: an empirical study. **Machine Learning**, Springer Science and Business Media LLC, v. 84, n. 1-2, p. 109–136, dez. 2010. ISSN 1573-0565.
- SMAGULA, S. F. et al. Moderators of response to cognitive behavior therapy for major depression in patients with heart failure. **Psychosomatic Medicine**, Ovid Technologies (Wolters Kluwer Health), v. 81, n. 6, p. 506–512, maio 2019. ISSN 0033-3174.
- STIGLIC, G. et al. Interpretability of machine learning-based prediction models in healthcare. **WIREs Data Mining and Knowledge Discovery**, Wiley, v. 10, n. 5, jun. 2020. ISSN 1942-4795.
- TAUBITZ, F.-S.; BUDENBENDER, B.; ALPERS, G. W. What the future holds: Machine learning to predict success in psychotherapy. **Behaviour Research and Therapy**, Elsevier BV, v. 156, p. 104116, set. 2022. ISSN 0005-7967.
- THEOBALD, O. **Machine Learning for Absolute Beginners: A Plain English Introduction**. 3. ed. [S.l.]: Scatterplot Press, 2021.
- WOODWORTH, R. J. et al. Web-based positive psychology interventions: A reexamination of effectiveness. **Journal of Clinical Psychology**, v. 73, n. 3, p. 218–232, 2017.
- WU, T. et al. A brief overview of chatgpt: The history, status quo and potential future development. **IEEE/CAA Journal of Automatica Sinica**, v. 10, n. 5, p. 1122–1136, 2023.



Pontifícia Universidade Católica do Rio Grande do Sul
Pró-Reitoria de Graduação e Educação Continuada
Av. Ipiranga, 6681 - Prédio 1 - 3º. andar
Porto Alegre - RS - Brasil
Fone: (51) 3320-3500 - Fax: (51) 3339-1564
E-mail: prograd@pucrs.br
Site: www.pucrs.br