

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 [4]. 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 [10]. Especula-se que um atendimento personalizado a nível individual possa melhorar a taxa de resposta ao tratamento [15, 14].

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 [23, 1]. 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 [24].

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 [6]. 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 [16]. Estudos sobre a aplicação de modelos de aprendizagem de máquina na previsão do desfecho de tratamentos em saúde mental apresentam resultados promissores [6], 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 [7, 17].

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 [3], 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 [2, 25]. 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 [25].

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 [2]. 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 [25].

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 [11] 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 [19]. 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 [25, 13].

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 [25, 2].

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 [25, 2]. 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 [3].

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 [25, 2]. Um exemplo para a aplicação deste tipo de aprendizagem é investigar os padrões de comorbidade em uma determinada população clínica [18].

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 [25, 2]. Um exemplo para a aplicação deste tipo de aprendizagem é auxiliar a tomada de decisões de tratamento em condições crônicas como a esquizofrenia [22].

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 [2, 13]. 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 [2].

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 [2, 13]. 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 [2]. Esse tipo de modelo é associado a ferramentas de inteligência artificial generativa como o Chat GTP [27].

2.3 A construção de uma aplicação de machine learning

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 [9].

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 [5].

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 [9].

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 [9].

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 [5].

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 [9]. Um exemplo de framework é o scikit-learn [20] para a linguagem de programação Python [8].

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 [5].

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 [9]. Modelos de aprendizagem supervisionada que realizam tarefas de regressão são avaliados com medidas de erro como o erro quadrático médio [5]. Modelos de aprendizagem supervisionada para tarefas de classificação utilizam métricas de acurácia preditiva como a área sob a curva ROC [9].

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 [12].

3 Construção de um exemplo

Introdução à construção do exemplo.

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 [26], uma tentativa de replicar os resultados obtidos em um trabalho anterior conduzido por Seligman e colaboradores [21].

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%) [26, 3].

Intervenção 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 [26].

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 [26].

Desfechos A.

3.2 Plano de análise de dados

A.

3.3 Resultados e discussão

A.

4 Conclusão

Texto da conclusão.

References

- [1] Margaret S. Andover et al. “Moderators of treatment response to an intervention for nonsuicidal self-injury in young adults.” In: *Journal of Consulting and Clinical Psychology* 88.11 (Nov. 2020), pp. 1032–1038. ISSN: 0022-006X. DOI: 10.1037/ccp0000603. URL: <http://dx.doi.org/10.1037/ccp0000603>.
- [2] Qifang Bi et al. “What is Machine Learning? A Primer for the Epidemiologist”. In: *American Journal of Epidemiology* (Oct. 2019). ISSN: 1476-6256. DOI: 10.1093/aje/kwz189. URL: <http://dx.doi.org/10.1093/aje/kwz189>.
- [3] Amanda C. Collins et al. “Predicting individual response to a web-based positive psychology intervention: a machine learning approach”. In: *The Journal of Positive Psychology* 19.4 (Sept. 2023), pp. 675–685. ISSN: 1743-9779. DOI: 10.1080/17439760.2023.2254743. URL: <http://dx.doi.org/10.1080/17439760.2023.2254743>.
- [4] Pim Cuijpers et al. “The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis”. In: *Acta Psychiatrica Scandinavica* 144.3 (July 2021), pp. 288–299. ISSN: 1600-0447. DOI: 10.1111/acps.13335. URL: <http://dx.doi.org/10.1111/acps.13335>.
- [5] Jaime Delgadillo. “Machine learning: A primer for psychotherapy researchers”. In: *Psychotherapy Research* 31.1 (Dec. 2020), pp. 1–4. ISSN: 1468-4381. DOI: 10.1080/10503307.2020.1859638. URL: <http://dx.doi.org/10.1080/10503307.2020.1859638>.
- [6] Dominic B. Dwyer, Peter Falkai, and Nikolaos Koutsouleris. “Machine Learning Approaches for Clinical Psychology and Psychiatry”. In: *Annual Review of Clinical Psychology* 14.1 (May 2018), pp. 91–118. ISSN: 1548-5951. DOI: 10.1146/annurev-clinpsy-032816-045037. URL: <http://dx.doi.org/10.1146/annurev-clinpsy-032816-045037>.
- [7] Conselho Nacional de Educação. “Resolução CNE/CES nº 1, de 11 de outubro de 2023”. In: *Diário Oficial União* (2023). URL: <https://www.in.gov.br/web/dou/-/resolucao-cne/ces-n-1-de-11-de-outubro-de-2023-518120795> (visited on 06/22/2024).
- [8] Python Software Foundation. *python.org*. <https://www.python.org/>. [Accessed 26-06-2024].
- [9] Joe G. Greener et al. “A guide to machine learning for biologists”. In: *Nature Reviews Molecular Cell Biology* 23.1 (Sept. 2021), pp. 40–55. ISSN: 1471-0080. DOI: 10.1038/s41580-021-00407-0. URL: <http://dx.doi.org/10.1038/s41580-021-00407-0>.

- [10] Stefan G. Hofmann et al. “The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses”. In: *Cognitive Therapy and Research* 36.5 (July 2012), pp. 427–440. ISSN: 1573-2819. DOI: 10.1007/s10608-012-9476-1. URL: <http://dx.doi.org/10.1007/s10608-012-9476-1>.
- [11] Cláudio S Hutz. *Avaliação Psicológica da Inteligência e da Personalidade*. Artmed, Apr. 2018.
- [12] Samreen Naeem et al. “An Unsupervised Machine Learning Algorithms: Comprehensive Review”. In: *International Journal of Computing and Digital Systems* 13.1 (Apr. 2023), pp. 911–921. ISSN: 2210-142X. DOI: 10.12785/ijcds/130172. URL: <http://dx.doi.org/10.12785/ijcds/130172>.
- [13] Andrew Ng and Michael Jordan. “On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes”. In: *Advances in Neural Information Processing Systems*. Ed. by T. Dietterich, S. Becker, and Z. Ghahramani. Vol. 14. MIT Press, 2001. URL: https://proceedings.neurips.cc/paper_files/paper/2001/file/7b7a53e239400a13bd6be6c91c4f6c4e-Paper.pdf.
- [14] John C. Norcross and Michael J. Lambert. “Psychotherapy relationships that work III.” In: *Psychotherapy* 55.4 (Dec. 2018), pp. 303–315. ISSN: 0033-3204. DOI: 10.1037/pst0000193. URL: <http://dx.doi.org/10.1037/pst0000193>.
- [15] John C. Norcross and Bruce E. Wampold. “What works for whom: Tailoring psychotherapy to the person”. In: *Journal of Clinical Psychology* 67.2 (Nov. 2010), pp. 127–132. ISSN: 1097-4679. DOI: 10.1002/jclp.20764. URL: <http://dx.doi.org/10.1002/jclp.20764>.
- [16] Jan A. Roth et al. “Introduction to Machine Learning in Digital Healthcare Epidemiology”. In: *Infection Control and Hospital Epidemiology* 39.12 (Nov. 2018), pp. 1457–1462. ISSN: 1559-6834. DOI: 10.1017/ice.2018.265. URL: <http://dx.doi.org/10.1017/ice.2018.265>.
- [17] Caio Rudá, Denise Coutinho, and Naomar de Almeida Filho. “Formação em psicologia: uma análise curricular de cursos de graduação no Brasil”. In: *Revista e-Curriculum* 17.2 (June 2019), pp. 419–440. ISSN: 1809-3876.
- [18] Marina Sánchez-Rico and Jesús M Alvarado. “A machine learning approach for studying the comorbidities of complex diagnoses”. en. In: *Behav. Sci. (Basel)* 9.12 (Nov. 2019), p. 122.
- [19] Carlos H Sancineto. *Bateria Fatorial da Personalidade - BFP: manual técnico*. Casa do Psicólogo, 2015.
- [20] *scikit-learn: machine learning in Python*. <https://scikit-learn.org/>. [Accessed 26-06-2024].
- [21] Martin E. P. Seligman et al. “Positive Psychology Progress: Empirical Validation of Interventions.” In: *American Psychologist* 60.5 (July 2005), pp. 410–421. ISSN: 0003-066X. DOI: 10.1037/0003-066x.60.5.410. URL: <http://dx.doi.org/10.1037/0003-066X.60.5.410>.

- [22] Susan M. Shortreed et al. “Informing sequential clinical decision-making through reinforcement learning: an empirical study”. In: *Machine Learning* 84.1-2 (Dec. 2010), pp. 109–136. ISSN: 1573-0565. DOI: 10.1007/s10994-010-5229-0. URL: <http://dx.doi.org/10.1007/s10994-010-5229-0>.
- [23] Stephen F. Smagula et al. “Moderators of Response to Cognitive Behavior Therapy for Major Depression in Patients With Heart Failure”. In: *Psychosomatic Medicine* 81.6 (May 2019), pp. 506–512. ISSN: 0033-3174. DOI: 10.1097/psy.0000000000000712. URL: <http://dx.doi.org/10.1097/PSY.0000000000000712>.
- [24] Friedrich-Samuel Taubitz, Björn Büdenbender, and Georg W. Alpers. “What the future holds: Machine learning to predict success in psychotherapy”. In: *Behaviour Research and Therapy* 156 (Sept. 2022), p. 104116. ISSN: 0005-7967. DOI: 10.1016/j.brat.2022.104116. URL: <http://dx.doi.org/10.1016/j.brat.2022.104116>.
- [25] Oliver Theobald. *Machine learning for absolute beginners*. Scatterplot Press, Jan. 2021.
- [26] Rosalind J. Woodworth et al. “Web-Based Positive Psychology Interventions: A Reexamination of Effectiveness”. In: *Journal of Clinical Psychology* 73.3 (2017), pp. 218–232. DOI: <https://doi.org/10.1002/jclp.22328>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jclp.22328>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jclp.22328>.
- [27] Tianyu Wu et al. “A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development”. In: *IEEE/CAA Journal of Automatica Sinica* 10.5 (2023), pp. 1122–1136. DOI: 10.1109/JAS.2023.123618.