

Análise do impacto da COVID-19 no desenvolvimento do transtorno de estresse pós-traumático na população brasileira

Vítor C. Ferrer¹, Thyago A. Salgado², Ives C. Passos², Mariana Recamonde-Mendoza¹

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brazil

²Departamento de Psiquiatria – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brasil

{vcrferrer,mrmendoza}@inf.ufrgs.br, {ivescpl,antonelli.thyago}@gmail.com

Abstract. *This study analyzes the development of post-traumatic stress disorder (PTSD) cases during the COVID-19 pandemic in the Brazilian population. Classifiers were built using Support Vector Machine and Random Forest models, which were tuned through Bayesian optimization. The best-performing model achieved a recall of 0.578 ± 0.111 , precision of 0.265 ± 0.108 , AUROC of 0.724 ± 0.056 , and specificity of 0.871 ± 0.048 . The SHAP values were used in the interpretation of the model, which indicated that symptoms of depression, anxiety, and loneliness; media consumption; and the use of benzodiazepines and opioids were the most relevant attributes for PTSD incidence.*

Resumo. *Este estudo analisa o desenvolvimento de casos de transtorno de estresse pós-traumático (TEPT) durante a pandemia da COVID-19 na população brasileira. Foram construídos classificadores com os modelos Support Vector Machine e Random Forest, os quais foram ajustados por otimização Bayesiana. O modelo com melhor desempenho alcançou revocação de $0,578 \pm 0,111$, precisão de $0,265 \pm 0,108$, AUROC de $0,724 \pm 0,056$ e especificidade de $0,871 \pm 0,048$. Os valores SHAP foram utilizados na interpretação do modelo, que indicaram sintomas de depressão, ansiedade e solidão; consumo de mídia; e uso de benzodiazepínicos e opioides como atributos mais relevantes para a incidência de TEPT.*

1. Introdução

A COVID-19 teve seu primeiro caso registrado em dezembro de 2019, na cidade de Wuhan, na China. O novo coronavírus SARS-CoV-2, agente etiológico da doença, espalhou-se globalmente em um curto intervalo de tempo e, em 11 de março de 2020, a Organização Mundial da Saúde (OMS) declarou oficialmente o surto como uma pandemia. Nesse contexto, o transtorno de estresse pós-traumático (TEPT) destacou-se como uma das condições psicopatológicas cuja gravidade foi acentuada durante o período pandêmico.

Os acontecimentos da pandemia da COVID-19 podem ser compreendidos como eventos que atendem ao Critério A dos critérios diagnósticos do TEPT [Gouzman et al. 2022], conforme estabelecido pelo *Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition* (DSM-5). Esse critério define um evento traumático como a exposição a uma situação envolvendo ameaça ou ocorrência real de morte, lesão grave ou violência sexual, em pelo menos uma das seguintes formas: (1) vivenciar diretamente o evento; (2) testemunhar pessoalmente o evento ocorrido com outras pessoas; (3)

tomar conhecimento de que o evento aconteceu com um familiar ou amigo próximo; e (4) sofrer exposição repetida ou extrema a detalhes aversivos do evento.

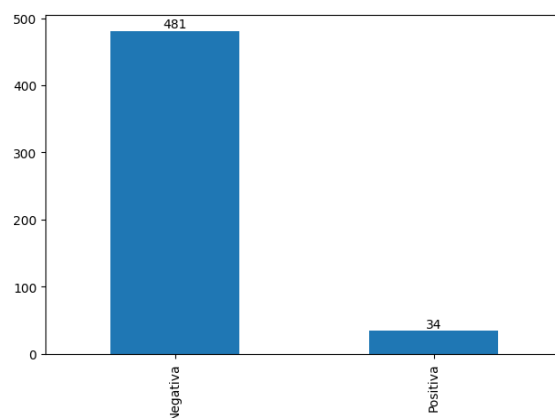
Para estender essa discussão, este trabalho objetiva analisar os fatores de risco associados à incidência de TEPT na população brasileira ao final da pandemia, bem como avaliar se o próprio período pandêmico pode ser caracterizado como um estressor traumático. As análises serão realizadas através de técnicas de aprendizado de máquina empregadas nas respostas de questionários online aplicados longitudinalmente ao longo da pandemia.

2. Metodologia

O objetivo do presente estudo é construir um modelo de aprendizado de máquina para prever casos de incidência de transtorno de estresse pós-traumático ao longo da pandemia da COVID-19 na população brasileira. O *dataset* utilizado, de natureza estruturada, foi coletado durante a pandemia da COVID-19 no Brasil pelo Dr. Thyago Antonelli e pelo Prof. Dr. Ives Cavalcante Passos, os quais conduziram um extenso estudo referente à saúde mental no período pandêmico [Antonelli-Salgado et al. 2021]. O conjunto foi segmentado em quatro ondas obtidas por meio de questionários online, sendo que a *baseline* foi obtida no início da pandemia, e as subsequentes foram coletadas após intervalos de 1, 6 e 24 meses em relação à primeira. Os questionários continham perguntas relativas às experiências dos participantes ao longo da pandemia, além da aplicação dos instrumentos psicométricos PHQ-9, PCL-5, GAD-7, R-UCLA, AUDIT-C e OSSS-3. Adicionalmente, o questionário da *baseline* incluía itens sobre características sociodemográficas dos indivíduos, como idade, orientação sexual e estado civil. Os dados foram pré-processados utilizando a linguagem *Python* com a biblioteca *pandas*.

A incidência corresponde a um dos quatro grupos nos quais os participantes foram categorizados, sendo os outros ausência, recuperação e persistência. Um participante é incidente quando não apresentava TEPT na onda 2, mas passou a apresentar na onda 4. A escolha da incidência como variável de predição visa possibilitar uma análise do impacto da exposição prolongada à pandemia na evolução do transtorno, buscando compreender os fatores de risco que podem ter sido catalisados por esse evento traumático prolongado. A quantidade de participantes diminuiu ao longo das diferentes ondas, sendo que a Figura 1 evidencia o alto desbalanceamento entre as classes negativa e positiva.

Figura 1. Distribuição da incidência de TEPT entre os participantes que responderam a segunda e a quarta onda dos questionários.



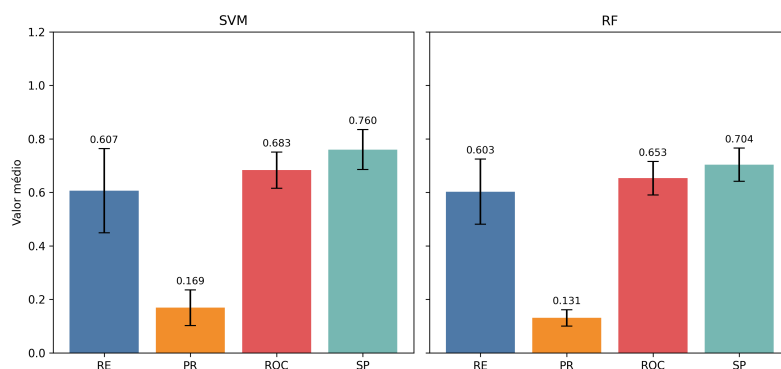
Os modelos de aprendizado de máquina escolhidos neste trabalho foram o *Support Vector Machine* (SVM) e *Random Forest* (RF). A escolha por esses dois algoritmos fundamenta-se em ambos possuírem desempenho satisfatório como preditores de TEPT em estudos realizados com populações gerais [Burton et al. 2024]. O treinamento desses modelos e o pré-processamento dos dados de entrada foram executados com algoritmos do *scikit-learn*. Foi construída uma pipeline de treinamento para a aplicação do método de validação cruzada aninhada, orquestrada pelo *framework Optuna*. Foram utilizadas $n = 10$ repetições para a validação externa e $n = 3$ repetições para a validação interna, com $k = 3$ folds em ambas. Por fim, o modelo com as melhores métricas foi analisado através dos valores SHAP.

3. Resultados e Discussão

Os resultados são apresentados em termos de média e desvio padrão para as seguintes métricas: revocação (RE), precisão (PR), área sob a curva ROC (AUROC) e especificidade (SP). A Figura 2 apresenta a média dessas métricas em múltiplas rodadas de treinamento para os dois modelos escolhidos. O SVM apresentou desempenho superior ao RF em todas as métricas analisadas, com destaque para a especificidade.

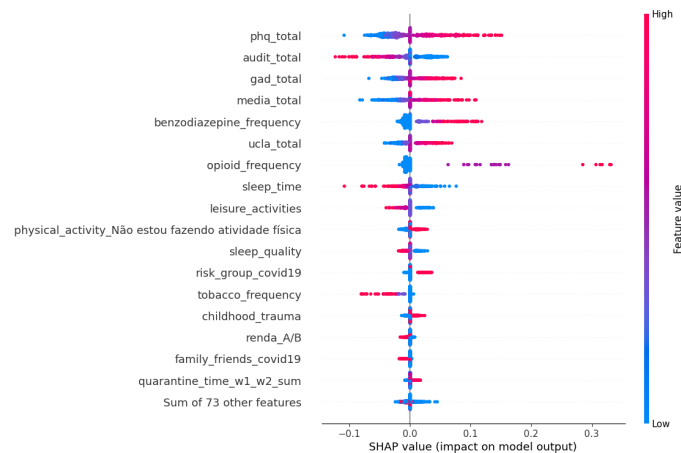
Destaca-se que ambos os modelos apresentaram a tendência a cometer muitos erros do tipo falso positivo em relação ao total de instâncias positivas, o que levou a uma baixa precisão. No entanto, métricas que avaliam outros aspectos do desempenho apresentaram resultados satisfatórios, indicando que a baixa precisão também pode ter sido influenciada pelo acentuado desbalanceamento do conjunto de dados.

Figura 2. Valores médios das métricas por tipo de modelo.



O modelo com melhor desempenho apresentou revocação de $0,578 \pm 0,111$, precisão de $0,265 \pm 0,108$, AUROC de $0,724 \pm 0,056$ e especificidade de $0,871 \pm 0,048$. A Figura 3 mostra os atributos com maior impacto nas predições, segundo os valores SHAP, em que cada ponto representa uma observação individual, enquanto a coloração indica a magnitude do valor do atributo (vermelho para valores altos e azul para baixos). Escores elevados nos instrumentos *PHQ*, *GAD* e *UCLA* — que avaliam sintomas de depressão, ansiedade e solidão — foram os principais fatores associados às predições positivas de *TEPT*. Também se destacaram o uso de benzodiazepínicos e opioides, este último com maior variação entre categorias. Em contraste, sintomas de alcoolismo (*AUDIT-C*) e consumo de tabaco apresentaram associação inversa com as predições do modelo.

Figura 3. Distribuição dos valores SHAP ordenados pelos atributos com as maiores contribuições.



4. Considerações finais

Os resultados obtidos indicam que a pandemia da COVID-19 constituiu um ambiente propício ao desenvolvimento de sintomas de TEPT. Nesse contexto, embora a pandemia, por si só, não se configure como um evento que atenda diretamente ao Critério A do DSM-5 — quando considerada apenas por suas circunstâncias específicas, como o *lockdown* e o distanciamento social — os achados deste estudo explicam a tendência da pandemia da COVID-19 apresentar uma maior incidência de TEPT em comparação a períodos não pandêmicos [Qiu et al. 2021].

Agradecimentos

Este estudo foi realizado com apoio da CAPES - Código de Financiamento 001, da FAPERGS [22/2551-0000390-7 (Projeto CIARS)] e do CNPq [308075/2021-8].

Referências

- Antonelli-Salgado, T., Monteiro, G. M. C., Marcon, G., Roza, T. H., Zimmerman, A., Hoffmann, M. S., Cao, B., Hauck, S., Brunoni, A. R., and Passos, I. C. (2021). Loneliness, but not social distancing, is associated with the incidence of suicidal ideation during the covid-19 outbreak: a longitudinal study. *Journal of Affective Disorders*, 290:52–60.
- Burton, M. S., Galatzer-Levy, I. R., Karstoft, K.-I., Baker, D. G., Lobo, S. L., Thompson, L., Gelaye, B., Saxbe, D., Nash, W. P., Fairholme, C. P., et al. (2024). Machine learning algorithms for predicting ptsd: a systematic review and meta-analysis. *Translational Psychiatry*, 14(1):172.
- Gouzman, J., Soskolne, V., and Dekel, R. (2022). Framing the meaning of covid-19 and the psychological responses to it: Insights gleaned from selected theoretical approaches. *Journal of Health Psychology*, 27(10):2331–2342.
- Qiu, D., Li, Y., Li, L., He, J., Ouyang, F., and Xiao, S.-Y. (2021). Prevalence of post-traumatic stress symptoms among people influenced by coronavirus disease 2019 outbreak: A meta-analysis. *European Psychiatry*, 64(1).