



**Visão sobre técnicas computacionais na detecção de depressão em texto**

**Insight into computational techniques in detecting depression in text**

**Información sobre técnicas computacionales para detectar la depresión en el texto**

Leidiane Rodrigues<sup>1</sup>, Deborah Fernandes<sup>2</sup>, Marilúcia Pereira do Lago<sup>3</sup>, Márcio Fernandes<sup>4</sup>, Fabrizzio Soares<sup>2</sup>, Kairo Silva<sup>1</sup>

<sup>1</sup> Mestrando(a) em Sistemas Inteligentes e Aplicações, Instituto de Informática, Universidade Federal de Goiás, Goiânia (GO), Brasil.

<sup>2</sup> Professor(a) Doutor(a), Instituto de Informática, Universidade Federal de Goiás, Goiânia (GO), Brasil.

<sup>3</sup> Professor(a) Doutor(a), Faculdade de Educação, Universidade Federal de Goiás, Goiânia (GO), Brasil.

<sup>4</sup> Professor Mestre, Campus de Ciências Exatas e Tecnológicas, Universidade Estadual de Goiás, Goiânia (GO), Brasil

Autor correspondente: Mestranda Leidiane Beatriz Passos Rodrigues  
*E-mail:* leidianebeatriz@discente.ufg.br

**Resumo**

**Objetivo:** Revisar a literatura sobre detecção de depressão em textos, focando em técnicas de aprendizado de máquina e processamento de linguagem natural.

**Método:** Análise de estudos que utilizam técnicas computacionais avançadas e dicionários de palavras indicativas de depressão, considerando a integração de métodos de aprendizado de máquina, processamento de linguagem natural, e recursos de saúde mental.

**Resultados:** A maioria dos trabalhos utiliza técnicas computacionais avançadas e dicionários específicos, mas há pouca integração da linguística e saúde mental nos modelos. Observou-se uma lacuna na incorporação do contexto cultural e regional da língua escrita.

**Conclusão:** Destaca-se a necessidade de incluir a linguística para considerar o contexto cultural e regional e aumentar o uso de recursos de saúde mental na identificação da depressão em textos, melhorando a precisão e eficácia das ferramentas de detecção.

**Descritores:** Revisão; Depressão; Computação.



### **Abstract**

**Objective:** Review the literature on depression detection in texts, focusing on machine learning and natural language processing techniques.

**Method:** Analysis of studies using advanced computational techniques and dictionaries of depression-indicative words, considering the integration of machine learning methods, natural language processing, and mental health resources.

**Results:** Most works use advanced computational techniques and specific dictionaries, but there is little integration of linguistics and mental health in the models. A gap in incorporating the cultural and regional context of written language was observed.

**Conclusion:** Emphasizes the need to include linguistics to consider the cultural and regional context and increase the use of mental health resources in identifying depression in texts, improving the precision and effectiveness of detection tools.

**Keywords:** Revision; Depression; Computing.

### **Resumen**

**Objetivo:** Revisar la literatura sobre la detección de la depresión en textos, centrándose en técnicas de aprendizaje automático y procesamiento de lenguaje natural.

**Método:** Análisis de estudios que utilizan técnicas computacionales avanzadas y diccionarios de palabras indicativas de depresión, considerando la integración de métodos de aprendizaje automático, procesamiento de lenguaje natural y recursos de salud mental.

**Resultados:** La mayoría de los trabajos utilizan técnicas computacionales avanzadas y diccionarios específicos, pero hay poca integración de la lingüística y la salud mental en los modelos. Se observó una brecha en la incorporación del contexto cultural y regional del lenguaje escrito.

**Conclusión:** Se destaca la necesidad de incluir la lingüística para considerar el contexto cultural y regional y aumentar el uso de recursos de salud mental en la identificación de la depresión en textos, mejorando la precisión y efectividad de las herramientas de detección.

**Descriptorios:** Revisión; Depresión; Informática.



## Introdução

A depressão é um transtorno mental que afeta milhões de pessoas em todo o mundo, manifestando-se de diferentes maneiras e intensidades. É uma condição complexa, caracterizada por sentimentos persistentes de tristeza, perda de interesse e prazer em atividades anteriormente apreciadas, além de uma variedade de sintomas físicos e emocionais (1). A prevalência da depressão varia significativamente entre diferentes faixas etárias, com crianças, adolescentes, adultos e idosos apresentando sintomas e desafios distintos. Nos jovens, por exemplo, a depressão pode se manifestar através de irritabilidade e problemas escolares, enquanto nos idosos pode ser confundida com sintomas de outras condições físicas ou cognitivas (2).

Diagnosticar a depressão é um desafio tanto para médicos quanto para psicólogos. Tradicionalmente, o diagnóstico é baseado em entrevistas clínicas, questionários e a observação dos comportamentos e relatos dos pacientes (3). No entanto, esses métodos podem ser subjetivos e dependem fortemente da capacidade do paciente em comunicar seus sentimentos e da experiência do profissional em interpretar esses sinais.

Com a crescente popularidade das redes sociais, muitas pessoas passaram a expressar suas emoções e sentimentos online, incluindo sintomas de depressão. Plataformas como Facebook, Twitter e Instagram são frequentemente utilizadas para compartilhar pensamentos e estados emocionais (4), proporcionando uma rica fonte de dados textuais que pode ser analisada para detectar sinais de depressão.

A identificação de depressão em textos por meio da psicologia e da medicina ainda é uma tarefa árdua. Tradicionalmente, profissionais de saúde mental analisam textos escritos por pacientes em contextos terapêuticos (5), mas a análise sistemática de grandes volumes de dados textuais nas redes sociais é limitada. Por outro lado, a computação oferece ferramentas avançadas para essa tarefa. Técnicas de inteligência artificial (IA) e processamento de linguagem natural (PLN) permitem a análise automatizada de grandes quantidades de texto, identificando padrões linguísticos e emocionais associados à depressão (6).

Atualmente, o diagnóstico da depressão na medicina e psicologia envolve avaliações clínicas detalhadas e questionários padronizados, como o Inventário de



Depressão de Beck (BDI) e a Escala de Depressão de Hamilton (HAM-D) (5). Esses métodos, embora eficazes, são complementados cada vez mais por abordagens computacionais. A IA e o PLN podem auxiliar na triagem inicial, identificando potenciais sinais de depressão em textos escritos pelos pacientes, permitindo um diagnóstico mais precoce e preciso.

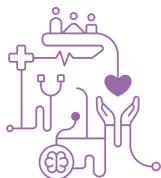
Uma questão crucial ao utilizar recursos computacionais para a detecção de depressão é o uso de léxicos<sup>1</sup> compostos por termos pré-definidos como depressivos. Esses léxicos funcionam como uma espécie de “gabarito”, verificando se determinados termos considerados depressivos estão presentes no texto analisado. Na análise automatizada em língua portuguesa, muitos léxicos utilizados são traduções literais de outras línguas, como o inglês (7). Alguns métodos empregam um léxico híbrido, combinando termos traduzidos com termos coletados em redes sociais na língua portuguesa. Outros utilizam dicionários compostos inteiramente por termos potencialmente depressivos coletados nas redes sociais, mas frequentemente sem considerar expressões culturais ou regionais específicas da língua (8).

A diferença entre usar um léxico traduzido para o um léxico coletado diretamente na língua de origem é significativa. Léxicos traduzidos podem não capturar nuances culturais e expressões idiomáticas específicas, resultando em interpretações imprecisas. Considerar a cultura e as expressões linguísticas da língua de origem pode ser relevante para compreender como as pessoas se expressam e, portanto, identificar sinais de depressão de maneira mais eficaz (9).

Dessa forma, visando contribuir para o estudo da detecção de depressão em textos em por meio da aplicação de técnicas avançadas de inteligência artificial e processamento de linguagem natural, incorporando abordagens que integrem a análise linguística e cultural, o presente trabalho realizou uma revisão da literatura nas principais bases de dados. O objetivo foi identificar os métodos computacionais mais utilizados, as principais redes sociais para coleta de dados, os recursos da área da saúde mental empregados e o uso da linguística como ferramenta de

---

<sup>1</sup> Léxico: No processamento de linguagem natural (PLN) e inteligência artificial (IA), um léxico é um conjunto organizado de palavras e expressões, contendo informações semânticas e gramaticais, usado para interpretar e gerar texto (10). Esses léxicos são essenciais para análises como a detecção de emoções em textos (11).



identificação de termos culturais e regionais na escrita para a detecção de depressão em textos.

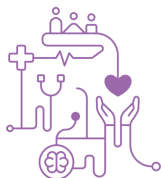
## Métodos

Esta revisão de literatura foi elaborada com o propósito de avaliar trabalhos sobre a detecção de depressão em textos escritos, utilizando técnicas avançadas de inteligência artificial e processamento de linguagem natural. Com o intuito de verificar e explorar esse tema foi realizada uma revisão exploratória focal utilizando o protocolo empregado quando se realiza uma revisão sistemática da literatura. Para esse trabalho, adotou-se o protocolo empregado por Kitchenham (12) para orientar a formulação da pergunta de pesquisa, a definição dos critérios de inclusão e exclusão, a busca e a seleção de estudos, a avaliação e análise dos resultados.

Para alcançar o objetivo do trabalho foram levantadas as seguintes questões de pesquisa:

- 1- Existem métodos de AM e PLN para identificar traços de depressão em texto?
- 2- Quais técnicas de PLN foram utilizadas para preparar e processar os dados no contexto de depressão?
- 3- Quais técnicas/métodos de PLN foram utilizados para identificar traços de depressão em textos?
- 4- Quais os tipos de dados textuais são utilizados em pesquisas de detecção automática de depressão em textos?
- 5- Existem trabalhos que abordam a construção de léxico para esse tema?
- 6- Existem trabalhos que abordam a adaptação de léxico em outras línguas para a língua de origem da análise dos dados?
- 7- Quais técnicas de AM foram empregadas para a construção de léxico?

A pesquisa eletrônica da produção científica foi realizada nas bases de dados IEEE, PubMed, Science Direct, Springer e OASIS. Para realizar as buscas nas bases de dados foram definidas *strings* de busca que contemplasse o escopo da computação, saúde mental e linguística, como é mostrado no Quadro 1 a seguir:



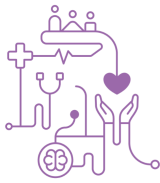
## Quadro 1 - Strings de busca.

( "machine learning" OR "aprendizado de máquina" OR "NLP", OR "processamento de linguagem natural" OR "social network" OR "twitter" OR "instagram" OR "reddit" OR "rede social online" OR "redes sociais") AND ("depressão" OR "ansiedade" OR "tristeza" OR "distímia" OR "criança ansiosa" OR "criança depressiva" OR "adoslescente" OR "adolescente ansioso" OR "adolescente depressivo" OR "depressão em texto" OR "ansiedade em texto" AND ("cultural" OR "aspectos culturais" OR "regionalismo" OR "aspectos regionais" OR "lugar cultural" OR "expressão cultural") )

Os artigos resultantes da string de busca foram exportados para uma planilha excel que foi utilizada para realizar a aplicação dos demais itens do protocolo. Os critérios para a extração de dados apresentados na tabela foram estabelecidos antes da leitura completa dos trabalhos selecionados. Esses critérios visavam orientar a identificação dos aspectos relevantes para abordar a questão de pesquisa, conforme detalhado no Quadro 2.

## Quadro 2 – Critérios de extração de dados e seus objetivos

Critério	Objetivo
Aplicação de recursos da computação	Artigos que utilizam alguma técnica computacional para identificar depressão como AM ou PLN.
Aplicação de métodos da área da saúde mental	Artigos que utilizam algum método ou protocolo de identificação de depressão da área da psicologia ou psiquiatria.
Uso da linguística	Artigos que utilizam a linguística para identificar traços psicolinguísticos na escrita.
Coleta de dados	Artigos que realizaram a coleta de dados para análise de depressão através de uma rede social ( Twitter, Instagram, Facebook, Reddit, etc)
Uso de léxico	Artigos que utilizaram algum léxico para análise de depressão em texto.



Como parte do protocolo proposto por Kitchenham (12), faz-se necessário definir os critérios de aceitação e exclusão, esses foram definidos conforme o explicitado no Quadro 3 e 4.

### **Quadro 3 – Critérios de aceitação**

CA1: O trabalho está relacionado a identificação de depressão;  
CA2: O trabalho utiliza técnicas de AM e/ou PLN na identificação de sentimentos em texto;  
CA3: O trabalho está utilizando base de dados de alguma rede social online ou entrevistas.  
CA4: O trabalho está utilizando um léxico;  
CA5: O trabalho está utilizando algum método da área da saúde mental para identificar depressão em texto.

### **Quadro 4 – Critérios de exclusão**

CE1: O trabalho não utiliza nenhuma técnica de AM ou PLN;  
CE2: O trabalho não tem como objetivo identificar depressão;  
CE3: O trabalho não usa nenhuma base de dados;  
CE4: O trabalho não utiliza base de dados textuais;  
CE5: O trabalho não está disponível por completo;  
CE6: O trabalho não apresenta nenhum referencial teórico;  
CE7: O trabalho está fora do escopo.

## **Resultados e Discussão**

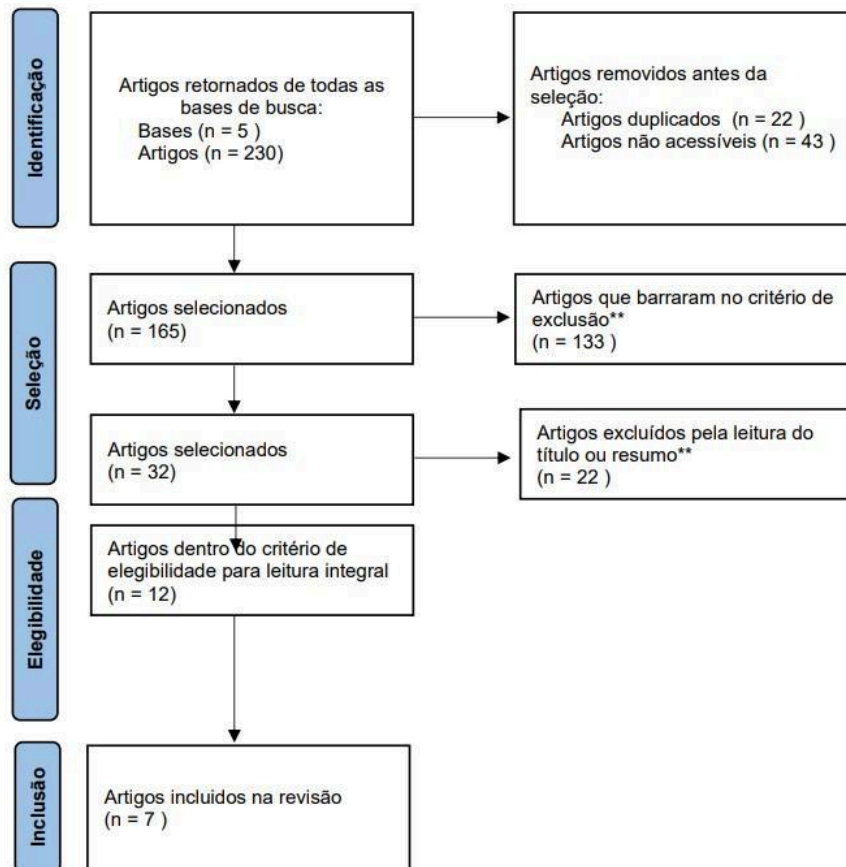
Com a aplicação das strings de busca nos motores de pesquisa, foram retornados 230 artigos possivelmente aptos para a revisão. Esse número foi reduzido para 165 devido à existência de 22 artigos duplicados e 43 não acessíveis de forma integral. Após a aplicação dos critérios de exclusão, conforme definido no Quadro 4, restaram 32 artigos a serem analisados por meio do título e da leitura do resumo. Após essa etapa, sobraram 12 artigos elegíveis para a leitura integral. No entanto, após a leitura integral, 5 artigos foram excluídos, pois investigavam a possível causa da depressão em vez de sua identificação, ou não utilizavam um léxico em sua análise, seja ele traduzido ou não para a língua portuguesa brasileira. Dessa forma, apenas 7 artigos foram incluídos nesta revisão. O fluxo para chegar a





esse número de artigos pode ser conferido no modelo PRISMA de revisão de literatura, na Figura 1.

**Figura 1 – Modelo PRISMA de revisão da literatura.**



Os 7 estudos desta revisão, conforme descritos no Quadro 5, fizeram o uso de aprendizado de máquina e processamento de linguagem natural utilizando um léxico como base de apoio para categorizar se um texto é depressivo ou não. Nesses trabalhos destacam-se os usos das técnicas *Support Vector Machines* (SVM), Processamento de Linguagem Natural (PLN), Naive Bayes (NB), Regressão Logística (RL), *Multilayer Perceptron* (MP) e Árvore de Decisão (AD) e *Convolutional Neural Network* (CNN) como as abordagens mais prevalentes nos estudos analisados. Essa predominância pode ser atribuída a eficácia comprovada dessas técnicas em lidar com desafios específicos relacionados a detecção de sinais de depressão em dados oriundos de plataformas de mídia social.





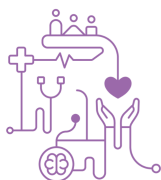
Outro aspecto interessante levantado na revisão é a realização de rotulação para o uso em técnicas de aprendizado de máquina supervisionados, é possível observar os métodos de rotulação adotados nos trabalhos, evidenciando a predominância da rotulação automática na maioria das análises. A prevalência da rotulação automática pode ser justificada pela sua eficiência em lidar com grandes volumes de dados de maneira rápida e escalável, enquanto a rotulação manual é frequentemente empregada quando se busca alta qualidade e precisão na anotação.

Outro ponto que podemos analisar é o uso das redes sociais empregadas nos trabalhos. As principais plataformas de rede social empregadas na coleta de dados, foram o Twitter, Facebook e Reddit. Vale ressaltar que o Twitter é notavelmente prevalente, por possuir, até 2023, uma API<sup>2</sup> de coleta robusta e aberta, o que facilitava a extração eficiente de dados. O Facebook, embora amplamente utilizado, apresenta restrições mais significativas em termos de acesso aos dados devido às políticas mais restritas de sua API e devido à privacidade dos perfis. Há redes onde a maioria dos usuários são de perfil aberto, nesse caso se destaca o Twitter, agora X. O Reddit, por sua vez, é uma plataforma popular para a obtenção de conteúdo textual, com uma API acessível. Outras fontes de dados mencionadas incluem o Instagram, para o qual as restrições de API podem variar. A diversidade nas escolhas das plataformas reflete a busca por variedade e abrangência nas fontes de dados, considerando tanto a disponibilidade quanto a acessibilidade das APIs para otimizar os processos de coleta e análise.

O uso de léxico não traduzido nos trabalhos também se mostrou prevalente, alguns estudos realizaram a tradução de léxicos do inglês para a língua materna do estudo para assim, realizar a detecção de sinais depressivos em texto. Foi possível notar também, que poucos estudos fazem o uso de recursos da área da saúde mental, como questionários e manuais de diagnósticos, para apoio na análise de um texto depressivo. Apenas um trabalho neste estudo, fez uso da linguística tentar para reconhecer termos depressivos em textos extraídos das mídias sociais como forma de expressar sentimentos em texto.

---

<sup>2</sup> Api Application Programming Interface (Interface de Programação de Aplicação), são mecanismos que permitem que dois componentes de software se comuniquem usando um conjunto de definições e protocolos (20).



### Quadro 5 – Estudos que detectam depressão em texto

Autor	Ano	Base de dados	T. C	T.S.M	Léxico traduzido?	Rotulação	Rede social
13	2022	IEE	PLN e BERT	-	Sim	Automática	Facebook
14	2020	Springer	SVM, LSTM e BERT	-	Não	Automática	Twitter
15	2022	Science Direct	CNN, LSTM e BERT	-	Não	Automática	Twitter e Reddit
16	2022	Science Direct	SVM e RF	TCC	Não	Automática	Entrevistas
17	2020	PubMed	PLN, RL, NB e SVM	-	Não	Automática	Facebook, Twitter, Reddit, Instagram
18	2021	OASIS	PLN, RL, SVM, AD e MP	PHQ-9	Sim	Automática	Twitter e Reddit
19	2018	IEE	NB e SVM	-	Não	Manual	Facebook

### Conclusão

Este estudo buscou identificar os métodos computacionais mais utilizados na detecção de depressão em textos, redes sociais para coleta de dados, léxicos traduzidos ou não, ferramentas ou métodos da área da saúde mental e o uso da linguística. Foi possível perceber que há vários estudos focados na identificação de sentimentos através de publicações em redes sociais, especificamente no Twitter, agora X, e no Reddit, devido à disponibilidade de APIs públicas e gratuitas para extração de dados na época.

Boa parte dos estudos optou por utilizar léxicos na mesma língua dos dados a serem analisados, sem traduzir termos potencialmente depressivos para a língua em que os dados seriam analisados. Quando se trata de um léxico traduzido, percebe-se uma redução na eficácia da identificação da depressão.

Observou-se pouco uso de recursos da saúde mental como ferramenta para a detecção de depressão em texto; a maioria dos trabalhos utilizou apenas léxicos de termos possivelmente depressivos para classificar os dados. A análise linguística revelou-se uma ferramenta valiosa para identificar termos relacionados à depressão na escrita, apresentando resultados significativos na melhora da identificação de usuários possivelmente depressivos.

Como próximos passos, recomendamos investigações mais aprofundadas sobre o uso de métodos e ferramentas da saúde mental em conjunto com técnicas



da computação, além do uso da linguística para identificar termos culturais e regionais. Outro ponto a ser estudado é o uso das LLMs (Large Language Models) nos trabalhos, tais como GPT e Gemini.

## Referências

1. de Souza RC. O que é psicologia. [Internet]. 2024 [citado 10 jan 2024]. Disponível em: [https://www.oasisbr.ibict.br/vufind/Record/UFAM-1\\_4b00bd2b922bad486fde3ef8829cd87b](https://www.oasisbr.ibict.br/vufind/Record/UFAM-1_4b00bd2b922bad486fde3ef8829cd87b)
2. Rice F, Riglin L, Lomax T, Souter E, Potter R, Smith D, Thapar A, Thapar A. Diferenças entre adolescentes e adultos nos perfis de sintomas de depressão maior. [Internet]. 2019 [citado 08 nov 2023]. doi: <https://doi.org/10.1016/j.jad.2018.09.015>.
3. Tolentino J, Schmidt S. Critérios do DSM-5 e gravidade da depressão: implicações para a prática clínica. *Fronteiras em Psiquiatria*. 2018;9 [citado 08 nov 2023]. doi: <https://doi.org/10.3389/fpsy.2018.00450>.
4. Vermeulen A, Vandebosch H, Heirman W. #Sorrindo, #desabafando ou ambos? Compartilhamento social de emoções por adolescentes nas redes sociais. *Computação. Zumbir. Comporte-se.* 2018;84:211-219. doi: <https://doi.org/10.1016/j.chb.2018.02.022>.
5. Trivedi M. Transtorno Depressivo Maior na Atenção Primária: Estratégias para Identificação. *The Journal of Clinical Psychiatry*. 2020;81(2) [citado 08 nov 2023]. doi: <https://doi.org/10.4088/jcp.ut17042br1c>.
6. Amanat A, Rizwan M, Javed A, Abdelhaq M, Alsaqour R, Pandya S, Uddin M. Aprendizado profundo para detecção de depressão a partir de dados textuais. *Eletrônicos*. 2022 [citado 08 nov 2023]. doi: <https://doi.org/10.3390/electronics11050676>.
7. World Health Organization. *Depression and Other Common Mental Disorders: Global Health Estimates*. World Health Organization; 2017.
8. Silva FA, Souza RS. Natural Language Processing for Social Media Text Analysis in Detecting Depression. *Journal of Medical Internet Research*. 2020;22(6)
9. Oliveira LM, Cunha AB. Cultural Adaptation and Validation of Depression Screening Tools in Brazil. *International Journal of Mental Health Systems*. 2019;13:45.
10. Manning CD, Raghavan P, Schütze H. *Introduction to Information Retrieval*. Cambridge University Press; 2008.
11. Jurafsky D, Martin JH. *Speech and Language Processing*. 3rd ed. Pearson; 2019.
12. Kitchenham B, Brereton P. A systematic review of systematic review process research in software engineering. *Information and Software Technology*. 2013;55(12):2049-2075.



13. Saravanan T, Jhaideep T, Bindu NH. Detecting depression using Hybrid models created using Google's BERT and Facebook's Fast Text Algorithms. Proceedings of the 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). 2022 Apr 28; Greater Noida, India. p. 415-421. doi: 10.1109/ICACITE53722.2022.9823581.
14. de Carvalho VF, Giacon B, Nascimento C, Nogueira BM. Aprendizado de Máquina para Identificação de Ideação Suicida no Twitter para a Língua Portuguesa. In: Cerri R, Prati RC, editores. Sistemas Inteligentes. BRACIS 2020. Notas de aula em Ciência da Computação. vol. 12319. Springer, Cham; 2020. p. 123-131. doi: 10.1007/978-3-030-61377-8\_37.
15. Malhotra A, Jindal R. Deep learning techniques for suicide and depression detection from online social media: A scoping review. Applied Soft Computing. 2022;130:109713. doi: 10.1016/j.asoc.2022.109713.
16. Vieira S, Liang X, Guiomar R, Mechelli A. Can we predict who will benefit from cognitive-behavioural therapy? A systematic review and meta-analysis of machine learning studies. Clinical Psychology Review. 2022;97:102193. doi: 10.1016/j.cpr.2022.102193.
17. Li G, Li B, Huang L, Hou S. Automatic Construction of a Depression-Domain Lexicon Based on Microblogs: Text Mining Study. JMIR Med Inform. 2020 Jun 23;8(6) doi: 10.2196/17650.
18. Lima GMdA. Detecção de indícios de depressão em textos curtos usando transferência de conhecimento. [Internet]. 2023 [citado 15 jan 2024]. Disponível em: [https://www.oasisbr.ibict.br/vufind/Record/UFAM-1\\_4b00bd2b922bad486fde3ef8829cd87b](https://www.oasisbr.ibict.br/vufind/Record/UFAM-1_4b00bd2b922bad486fde3ef8829cd87b)
19. Cha J, Kim S, Park E. A lexicon-based approach to examine depression detection in social media: the case of Twitter and university community. Humanit Soc Sci Commun. 2022;9(1):325. doi: 10.1057/s41599-022-01313-2.
20. O que é API. [Internet]. 2024 [citado 01 maio 2024]. Disponível em: <https://encurtador.com.br/nUCzg>