

Bem-te-vejo: Aplicação para análise e visualização de dados depressivos a partir de postagens do Reddit

Felipe Targino do Nascimento¹, Lucas Andrade dos Santos¹,
Pablo de Lima Marques Estrela¹, Ayrton Douglas Rodrigues Herculano¹, Damires
Yluska de Souza Fernandes¹, Alex Sandro da Cunha Rego¹

¹Instituto Federal da Paraíba (IFPB)

Av. 1º de Maio, 720 – Jaguaribe – João Pessoa – PB – Brasil

{targino.felipe,lucas.santos18,pablo.estrela,
ayrton.herculano}@academico.ifpb.edu.br, {damires,alex}@ifpb.edu.br

Abstract. *Depression is one of the leading causes of disability worldwide. During the Covid-19 pandemic, this scenario worsened, increasing the number of people affected by the disorder and boosting the use of social media. In this light, this work presents “Bem-te-vejo”, a web application developed to ease the analysis of depressive posts on Reddit. It also provides a predictive model, which is able to classify if a given post is likely to be depressive.*

Resumo. *A depressão é uma das principais causas de incapacidade em todo mundo. Durante a pandemia de Covid-19, esse cenário se intensificou, elevando o número de pessoas acometidas pelo transtorno e impulsionando o uso de redes sociais. Nesse contexto, este trabalho apresenta a “Bem-te-vejo”, uma aplicação web desenvolvida para facilitar a análise de postagens depressivas do Reddit. A aplicação provê também o uso de um classificador capaz de indicar se uma postagem tem teor depressivo.*

1. Introdução

A depressão é um transtorno mental crônico e recorrente, caracterizada por sentimentos como angústia, baixa autoestima e perda de interesse nas atividades do cotidiano (OMS, 2023). Segundo a Organização Mundial da Saúde (OMS), mais de 300 milhões de pessoas são afetadas globalmente por esse transtorno, sendo as mulheres as mais vulneráveis. Durante a pandemia de Covid-19, os casos de depressão aumentaram em 25% ao redor do mundo (OMS, 2023). Neste período, diante das restrições de circulação e do isolamento social, a utilização dos meios digitais foi intensificada, fazendo com que muitas pessoas recorressem às redes sociais para externar seus sentimentos e buscar, inclusive, apoio emocional (CHO et al., 2023). Uma das redes sociais que refletiram essa demanda foi a Reddit¹, um espaço virtual composto por subcomunidades que agrupam interações e mensagens de acordo com um tema, dentre os quais a depressão. Exemplos de subcomunidades inseridas nessa temática incluem "r/depression" e "r/mentalhealth".

Postagens e comentários compartilhados em redes sociais constituem uma fonte secundária de informações que, uma vez coletadas e tratadas usando técnicas de Processamento de Linguagem Natural (PLN) e de Análise de Sentimentos (AS), podem ser utilizadas para detectar conteúdo de cunho depressivo (CASELI; NUNES, 2024).

¹ <https://www.reddit.com/>

Extrair e analisar um volume considerável de postagens pode possibilitar a descoberta de padrões e tendências que seriam difíceis de serem identificados em um pequeno conjunto de dados (FRANÇA et al., 2014). Uma das abordagens para a análise de uma grande quantidade de postagens consiste na utilização de ferramentas de análise e visualização de dados, as quais podem contribuir significativamente para a compreensão e identificação de padrões e tendências no uso das redes sociais. Nesse contexto, este artigo apresenta a “Bem-te-veja”, uma aplicação web desenvolvida com o objetivo de auxiliar na análise de dados provenientes de postagens extraídas de redes sociais, com possível teor depressivo. A ferramenta possui uma interface interativa com filtros customizáveis e dinâmicos, que permite refinar a razão das análises conforme os critérios previamente definidos pelos usuários. A Bem-te-veja integra um classificador ajustado a partir do modelo pré-treinado DepreBERTBR, que analisa postagens de entrada e as classifica quanto ao nível de intensidade de depressão (ausente, moderada ou grave), conforme trabalho realizado por (HERCULANO et al., 2024).

O presente artigo está organizado da seguinte forma: A Seção 2 discute alguns trabalhos relacionados. A Seção 3 apresenta a aplicação Bem-te-veja. Por fim, a Seção 4 tece algumas considerações e indica trabalhos futuros.

2. Trabalhos Relacionados

A SoNDA (RAMOS et al., 2016) é uma aplicação que auxilia na análise e classificação manual de postagens do X² (anteriormente Twitter). A ferramenta emprega métodos de classificação que permitem agrupar os posts por palavras chaves e categorias previamente definidas pelos usuários. Embora esses métodos forneçam uma opção de filtragem inicial, a classificação final das postagens é realizada manualmente, exigindo uma leitura cuidadosa por parte do usuário anotador.

Seman (2021) desenvolveu o FYP (*Final Year Project*) *Depression Dashboard*, um painel interativo para monitorar sentimentos em postagens de cunho depressivo extraídas da plataforma X. A aplicação utiliza um modelo preditivo para classificar as postagens como positivas, negativas ou neutras, apresentando os resultados por meio de indicadores quantitativos, gráficos temporais e nuvem de palavras. Já o Emognizer (ANDRADE, 2020) é um aplicativo móvel que utiliza redes neurais para classificar emoções como raiva, medo e tristeza expressas em postagens do X e do Reddit. Sua interface exibe as análises de forma intuitiva, utilizando gráficos dinâmicos. Contudo, o aplicativo não realiza a classificação das postagens em relação à presença de algum transtorno mental, como a depressão.

A UnB Sense (SPERLING, 2019) é uma aplicação que utiliza um modelo preditivo para classificar perfis de usuários na plataforma X quanto a possíveis indícios de depressão, considerando padrões de uso e interação, como a frequência de postagens e a presença de palavras-chaves. Apesar de oferecer uma interface intuitiva e interativa, a aplicação apresenta limitações relacionadas à ausência de mecanismos de filtragem, o que pode dificultar a segmentação dos dados e execução de uma análise mais precisa dos dados. A ferramenta desenvolvida por Bernice Ziwei et al. (2019) realiza a classificação de postagens do X como depressivas ou não, utilizando as bibliotecas TextBlob e VADER (*Valence Aware Dictionary and sEntiment Reasoner*), aplicadas

² <https://x.com/>

diretamente nos textos. Entretanto, a ferramenta carece de opção de filtragem e painéis interativos de visualização, o que pode dificultar a interpretação dos resultados.

Como diferenciais do presente trabalho, a Bem-te-vejo oferece diversas opções de filtragem de dados e um mecanismo de classificação que utiliza um modelo preditivo treinado para identificar três possíveis classes de depressão (HERCULANO et al., 2024). Isso permite uma análise mais precisa ao considerar variações sutis da linguagem proferida nas postagens. A Bem-te-vejo provê uma interface interativa que, por exemplo, assiste o usuário na análise de ocorrências de termos do domínio da depressão, assim como na identificação da frequência de publicações sobre o tema.

3. Bem-te-vejo: aplicação para análise de dados

A Bem-te-vejo é uma aplicação web desenvolvida com o intuito de facilitar a compreensão, análise e interpretação de dados da rede social Reddit no domínio da Depressão. Sua interface permite realizar consultas customizadas utilizando mecanismos de filtragem temporal (e.g., dia da semana), palavras-chaves (termos), e conteúdo (e.g., título, texto e comentários). Por meio de gráficos, o usuário pode identificar comportamentos embutidos nos dados. A Bem-te-vejo introduz o uso de um classificador multiclasse ajustado a partir do modelo DepreBERTBR (HERCULANO et al., 2024), um modelo de linguagem pré-treinado com um grande volume de textos extraídos de postagens compartilhadas na rede social Reddit³. O classificador é capaz de atribuir a um dado texto de entrada um dos 3 (três) rótulos seguintes para depressão: ausente, moderado ou grave.

3.1 Arquitetura

A Figura 1 ilustra uma visão geral da arquitetura da aplicação Bem-te-vejo, subdividida em três camadas principais: (i) interface, responsável por prover ao usuário a interação com o sistema; (ii) serviços, que atuam no processamento das requisições da aplicação, como as consultas ao banco de dados e as solicitações de predição do modelo DepreBERTBR; e (iii) camada de dados, que gerencia o fluxo de requisições de armazenamento e recuperação de textos com um Sistema NoSQL.

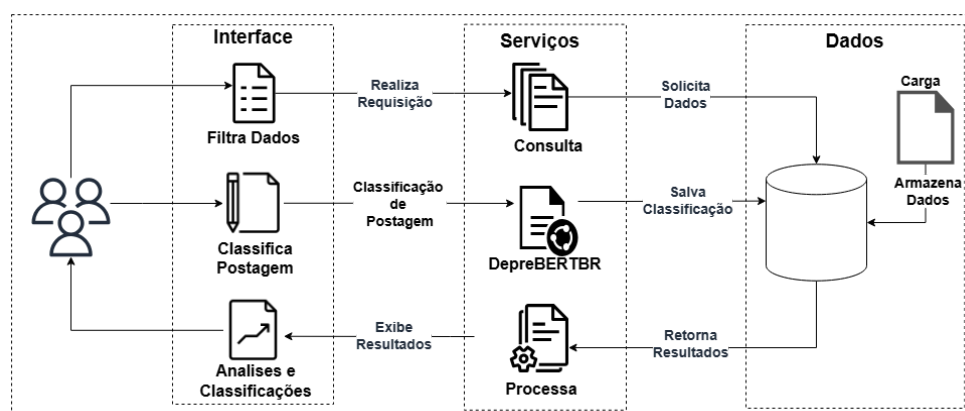


Figura 1 - Arquitetura da aplicação Bem-te-vejo

³ As postagens são extraídas de subcomunidades temáticas ao discurso depressivo.

3.2 Processo de Extração, Transformação e Carga

A preparação da base de dados utilizada pela aplicação Bem-te-vejo foi guiada por um processo de Extração, Transformação e Carga (*Extract, Transform, Load* - ETL), com o objetivo de lidar adequadamente com as tarefas de coleta, processamento e armazenamento de postagens oriundas do Reddit. A etapa de extração de postagens exigiu o desenvolvimento de uma ferramenta de Extração (ARAÚJO et al., 2024), concebida para coletar postagens do Reddit considerando as subcomunidades alvo. As postagens foram coletadas compreendendo o período de 16 de fevereiro de 2018 a 27 de agosto de 2023, seguindo as diretrizes de privacidade do Reddit, que permite que os dados sejam extraídos para utilização em meios acadêmicos ou pesquisas. Um estudo prévio identificou as subcomunidades de interesse relacionadas à saúde mental e as palavras-chave usadas para seleção de postagens (um texto é selecionado para coleta se apresentar a ocorrência de uma das palavras-chave). A relação inicial das palavras-chaves (e.g., “deprê”, “morrer”, “crises”, “suicídio”, “desespero”, “chorar” etc.) foi obtida a partir de uma revisão da literatura acerca da temática “detecção de depressão em conteúdo textual” (HERCULANO et al., 2022) e, posteriormente, validada por especialistas do domínio (médicos psiquiatras). O resultado culminou em um quantitativo de 200.030 exemplos de dados. Na etapa de transformação, foram padronizadas as informações de data e hora das postagens e acrescentadas colunas informativas sobre o dia da semana e período do dia da postagem. Procedeu-se, então, à carga dos dados utilizando o Sistema NoSQL Elasticsearch, adequado à análise e busca de dados textuais. Ao final, um total de 5.915 postagens e seus respectivos comentários produziram uma base com 126.060 textos.

3.3 Resultados e Análises

A Bem-te-vejo é composta por 5 (cinco) unidades de informação/interação com o usuário, sendo as principais relacionadas à exibição de indicadores, preparação de consulta e classificação de postagens. A página de Indicadores (Figura 2a) ilustra um panorama geral da distribuição das postagens de cunho depressivo, permitindo que os usuários consigam identificar padrões e tendências relevantes, como a variação de postagens ao longo dos dias, meses e anos. Essa página inclui visualizações gráficas que retratam a frequência de postagens ao longo dos anos, segmentada por dias da semana e por períodos do dia.

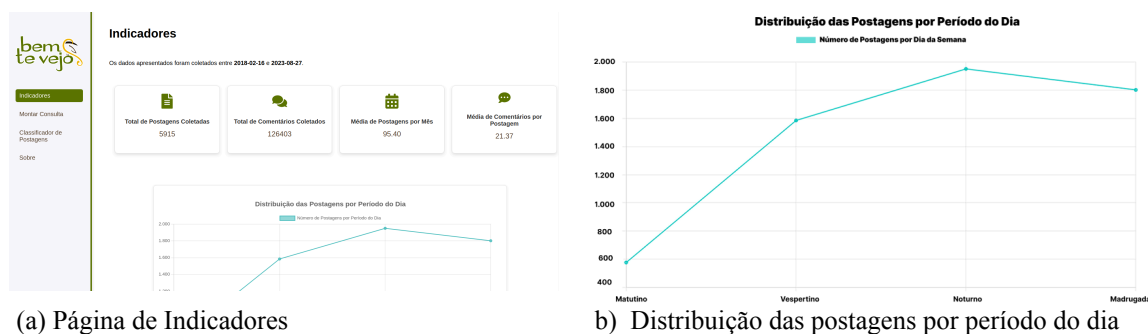
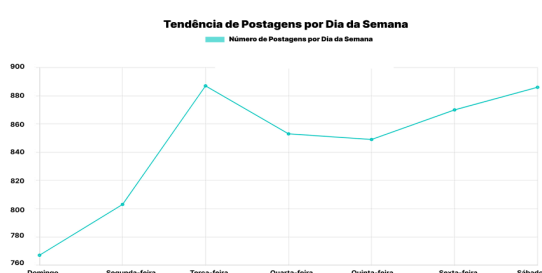
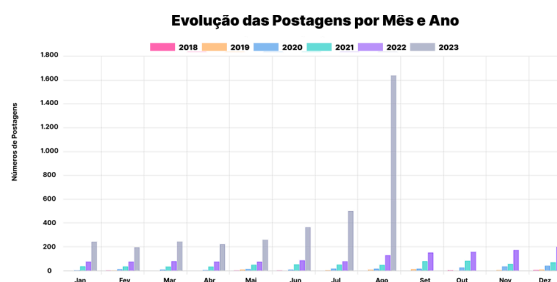


Figura 2 - Visão da Página de Indicadores.

A análise de distribuição por período do dia (Figura 2b) revela que a maioria das interações dos usuários no Reddit ocorreu durante a noite (1.951 postagens) e na madrugada (1.802 postagens), representando, juntos, um total de 63,44% do total de postagens. No que concerne à distribuição por dia da semana (Figura 3a), observa-se um pico de postagens às terças-feiras (887 postagens), seguido de uma leve tendência de queda nos dias subsequentes. O domingo obteve o menor volume de postagens (767).



(a) Tendência de Postagens por Dia da Semana



(b) Evolução das Postagens por Mês e Ano

Figura 3 - Gráficos de tendência e evolução das postagens.

A análise anual (Figura 3b) evidencia um crescimento progressivo no volume de postagens ao longo dos anos e meses, com destaque para o ano de 2023. Esse aumento pode indicar que houve tanto um aumento no número de usuários na utilização da rede, quanto uma intensificação na manifestação de sinais depressivos por parte desses usuários.

A página de construção de consulta (Figura 4a) permite que o usuário selecione postagens com base em filtros personalizados, como intervalo de datas e palavras-chave, permitindo então investigar determinados comportamentos, padrões e tendências específicas nas postagens. Os resultados das consultas podem ser visualizados de duas formas distintas: textual ou graficamente. Na visualização textual, o usuário pode analisar subjetivamente o conteúdo textual das postagens/comentários retornados. Já na segunda visualização (Figura 4b), são gerados gráficos que permitem agrupar o quantitativo de postagens por mês/ano e visualizar os termos mais utilizados nas postagens, o que favorece a identificação de tendências, padrões e variações de comportamento discursivo dos usuários ao longo do tempo.

Montar Consulta

Use os filtros abaixo para visualizar as postagens de acordo com os critérios selecionados.

Os dados apresentados foram coletados entre 2018-02-16 e 2023-08-27.

Formato de Saída: ☐ Textual ☐ Gráfico

Data Inicial: Data Final:

Período do Dia: ☐ Madrugada ☒ Matutino ☐ Vespertino ☐ Noturno

Dia da Semana: ☒ Segunda ☐ Terça ☐ Quarta ☐ Quinta ☐ Sexta ☐ Sábado ☐ Domingo

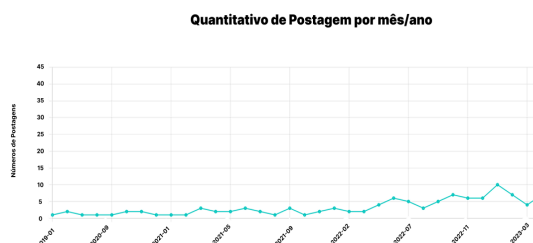
Filtrar por: ☒ Título ☐ Texto ☐ Comentários

Conteúdo do Filtro:

Total de Registros: 51 registros encontrados

Título da Postagem:

(a) Página de elaboração de consulta

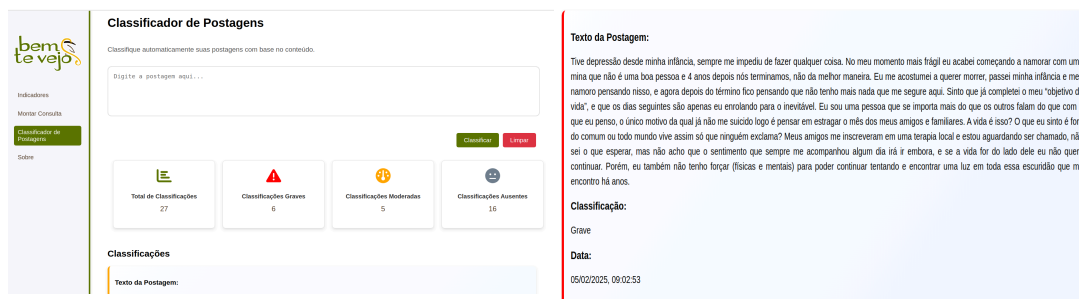


(b) Frequência de publicação de postagens

Figura 4 - Montagem de consultas

O modelo preditivo DepreBERTBR atua de forma indireta quando o usuário interage com a página de classificação de postagens (Figura 5a). Nessa funcionalidade,

o usuário pode inserir um texto livre e obter a sua respectiva classificação quanto ao nível de indício de depressão observado no texto da mensagem: ausente, moderado ou grave. O resultado da classificação é exibido com um destaque visual na vertical na qual a cor cinza denota um texto ausente de conteúdo depressivo, amarelo para um conteúdo que sugere uma conotação depressiva moderada e vermelho para aqueles com uma conotação depressiva de intensidade grave (Figura 5b).



(a) Classificador de postagem

(b) Postagem classificada

Figura 5 - Classificação de postagem

A Bem-te-veja se destaca como uma ferramenta inovadora e acessível, voltada à análise de dados de postagens de redes sociais relacionadas à depressão. Ao integrar visualizações dinâmicas, a aplicação permite que os usuários consigam identificar padrões e tendências expressos na rede social mais facilmente e com uma maior precisão. Ao permitir que os usuários obtenham a classificação automática de uma dada postagem, a aplicação viabiliza análises mais aprofundadas sobre o teor da postagem, inclusive por parte de profissionais de saúde. Essa combinação de mecanismos de visualização, filtragem e classificação torna a Bem-te-veja uma ferramenta relevante, com potencial de aplicação tanto em pesquisas acadêmicas como na área da saúde. A aplicação está disponível em <http://bemteveja.ifpbapps.com>.

4. Considerações e trabalhos futuros

Este trabalho apresentou a aplicação Bem-te-veja, uma solução desenvolvida para sumarizar dados em diferentes visões para a realização de análises, identificação de padrões e tendências a partir de postagens extraídas da rede social Reddit. A aplicação oferece uma interface intuitiva para explorar dados utilizando combinações de diferentes elementos de consulta. Embora o desenvolvimento da Bem-te-veja tenha como motivação inicial a investigação, análise e identificação de conteúdo depressivo, pode-se ajustá-la facilmente para processar dados a respeito de outros transtornos mentais em uma determinada janela de tempo, como também de outra rede social como o X. Como trabalhos futuros, destaca-se a necessidade de realizar nova extração e carga de dados para um período de tempo mais recente, bem como implementar uma etapa de avaliação sistemática da ferramenta, por meio da realização de testes e coleta de feedbacks de usuários.

Referências

ANDRADE, V. EMOGNIZER: Aplicação baseada em inteligência artificial para análise emocional de redes sociais. 2020. Trabalho de Conclusão de Curso. Universidade

- Regional Integrada do Alto Uruguai e das Missões, Campus de Erechim, 2020.
Disponível em:
<https://repositorio.uricer.edu.br/items/6a47379f-4360-4756-9d40-43ca81cd1aec>.
Acesso em: 2 dez 2024.
- ARAÚJO, A. et al. Uma ferramenta para extração de dados de redes sociais. João Pessoa: SECT 2024, Eixo Temático: Ciências Computacionais. Disponível em: https://www.ifpb.edu.br/joaopessoa/pesquisa/nova-pasta/xix-semana-de-educacao-ciencia-e-tecnologia-2024/anais_sect_2024_ifpb-02_05_2025.pdf. Acesso em: 2 dez 2024.
- CASELI, H.M.; Nunes, M.G.V. (org.) Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português. 3 ed. BPLN, 2024. Disponível em: <https://brasileiraspln.com/livro-pln/3a-edicao>
- CHO, Hichang et al. The bright and dark sides of social media use during COVID-19 lockdown: Contrasting social media effects through social liability vs. social support. *Computers in Human Behavior*, v. 146, p. 107795, 2023.
- CHOUDHURY, M.; GAMON, Michael; COUNTS, Scott; HORVITZ, Eric. Predicting Depression via Social Media. In: PROC. Of the 17th INTERNATIONAL AAAI CONFERENCE ON WEBLOGS AND SOCIAL MEDIA, 2013.
- COSTA, F. Visualização de dados e sua importância na era do Big Data. 2017. TCC. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017. Disponível em: <https://lume.ufrgs.br/handle/10183/169109>. Acesso em: 2 dez 2024.
- FRANÇA, T. C., et al. (2014). Big Social Data: Princípios sobre Coleta, Tratamento e Análise de Dados Sociais. In *Tópicos em Gerenciamento de Dados e Informações* (pp. 1-28). SBC. ISBN 978-85-7669-290-4.
- HERCULANO, A. et al. DepreBERTBR: Um Modelo de Linguagem Pré-treinado para o Domínio da Depressão no Idioma Português Brasileiro. In: *Proceedings of the 39th Brazilian Symposium on Databases*, outubro 2024, Florianópolis, SC. Anais [...].
- HERCULANO, A. D. R. ; GOMES, G. R. S. ; Cunha, A. ; SOUZA, Damires . Detecting Signs of Mental Disorders on Social Networks: a Systematic Literature Review. In: *DATA ANALYTICS 2022 (The Eleventh International Conference on Data Analytics)*, 2022, Valencia. *Proceedings of the The Eleventh International Conference on Data Analytics (Data Analytics, 2022)*. Valencia: IARIA, 2022. v. 1. p. 55-61.
- Organização Mundial de Saúde (OMS). Folha informativa – Depressão. Pandemia de COVID-19, 2022. Disponível em: <https://www.paho.org/pt/topicos/depressao>. Acesso em: 11 mai. 2023.
- RAMOS, V.; SMANIOTTO, C. Aplicação Web SoNDA: Social Network Data Analysis. In: *Pesquisa nas Redes*. 2015.
- SEMAN, F. Dashboard for Monitoring Depression Amongst Twitter Users Using Sentiment Analysis. 2021. Dissertação de Mestrado. Universiti Teknologi PETRONAS, Perak, Malásia, 2021. Disponível em: <https://utpedia.utp.edu.my/id/eprint/24091/>. Acesso em: 2 dez 2024.
- SPERLING, O. UnB Sense: a web application to probe for signs of depression from user profiles on social media. 2019. Trabalho de Conclusão de Curso. Universidade de Brasília, Instituto de Ciências Exatas, 2019.
- ZIWEI, B. YEOW; CHUA, H. An Application for Classifying Depression in Tweets. In: *International Conference on Computing and Big Data (ICCBD)*, Taichung, Taiwan, 2019. ACM, 2019.