

Detecção de Traços de Depressão em Textos na Língua Portuguesa Considerando Aspectos Culturais e Regionais do Brasil com o uso de LLMs

Leidiane Rodrigues^{1,3}, Deborah Fernandes^{1,3}, Marilúcia Lago^{2,3}

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

²Faculdade de Educação – Universidade Federal de Goiás (UFG)
Goiânia, GO – Brasil

³AKCIT - Centro de Competência EMBRAPII em Tecnologias Imersivas
Universidade Federal de Goiás (UFG) – Goiânia, GO – Brasil

leidianebeatriz@egresso.ufg.br, {marilucia.lago, deborah.fernandes}@ufg.br

Abstract. This study investigates the identification of depression traits in messages from the social network Reddit, focusing on young people and adolescents. To this end, two lexicons were compared: one with terms in Portuguese without cultural expressions and another culturally enriched lexicon, constructed from Brazilian song lyrics. Using the GPT 3.4 and Flash 1.5 models, we analyzed messages labeled as depressive and non-depressive, classified by a mental health specialist. The results show that the cultural adaptation of the lexicon allowed a more sensitive and accurate analysis, suggesting the importance of regional lexicons to improve the detection of signs of depression in specific cultural contexts.

Resumo. Este trabalho investiga a identificação de traços de depressão em mensagens da rede social Reddit, com foco em jovens e adolescentes. Para isso, foram comparados dois léxicos: um com termos em português sem expressões culturais e outro léxico enriquecido culturalmente, construído a partir de letras de músicas brasileiras. Utilizando os modelos GPT 3.4 e Flash 1.5, analisamos mensagens rotuladas como depressivas e não depressivas, classificadas por uma especialista em saúde mental. Os resultados mostram que a adaptação cultural do léxico permitiu uma análise mais sensível e precisa, sugerindo a importância de léxicos regionais para aprimorar a detecção de sinais de depressão em contextos culturais específicos.

1. Introdução

A depressão é um transtorno mental que impacta profundamente a vida dos indivíduos, afetando sua qualidade de vida, produtividade e bem-estar emocional. A detecção precoce da depressão é um desafio crítico para a área da medicina e da saúde mental, especialmente entre jovens e adolescentes, que muitas vezes apresentam dificuldades em expressar seus sentimentos ou procurar ajuda profissional. Nos últimos anos, ferramentas de Inteligência Artificial têm sido utilizadas para analisar sinais depressivos em textos, possibilitando uma triagem inicial que pode auxiliar especialistas no diagnóstico e

intervenção precoce. No entanto, grande parte dessas abordagens ignora aspectos culturais e regionais, o que pode comprometer a precisão da análise. Este estudo propõe a construção de um léxico adaptado ao contexto brasileiro e sua aplicação em Modelos de Linguagem de Grande Escala (LLMs) para a detecção de depressão em textos na língua portuguesa, contribuindo para um rastreamento mais sensível e adequado à realidade do país.

2. Objetivos e Contribuições

Este trabalho tem como:

- **Objetivo Geral:** Desenvolver um léxico voltado à identificação de traços depressivos em textos escritos em português brasileiro, incorporando elementos culturais e regionais.
- **Objetivos Específicos:**
 - Estabelecer uma abordagem interdisciplinar entre computação, saúde mental e linguística;
 - Definir critérios para a seleção de termos depressivos considerando variações culturais;
 - Construir e validar um léxico enriquecido que melhore a precisão dos modelos de IA à identificação de traços depressivos em textos escritos em português brasileiro.

As principais contribuições desta pesquisa são a construção de um léxico adaptado ao contexto linguístico brasileiro e a melhoria na precisão dos modelos de detecção de depressão.

3. Metodologia

A solução proposta envolveu um experimento dividido em duas etapas: experimento inicial e experimento final, explorando a extração de termos culturais e regionais a partir de letras de músicas brasileiras para detectar sinais de depressão em textos.

3.1. Experimento Inicial

O experimento inicial foi conduzido para avaliar a viabilidade da detecção de traços de depressão em textos utilizando um léxico adaptado ao contexto cultural e regional brasileiro. Ele seguiu três etapas principais:

- **Coleta de letras de músicas:** Foram selecionadas músicas dos gêneros brega, funk e MPB, com base na popularidade dos artistas, determinada por meio da plataforma de streaming de músicas *Spotify* [Spotify 2025], e na presença de expressões associadas à tristeza e depressão. As letras foram armazenadas integralmente em um banco de dados, sem remoção de estrofes repetidas. O processo contou com a orientação de um especialista em saúde mental para evitar vieses na seleção;
- **Extração de Termos:** Utilizando o ChatGPT 3.5 [FIA 2024], os termos culturais e regionais criando uma lista de expressões culturais e regionais associadas à tristeza e depressão a partir das letras de músicas. O prompt [Canaltech 2024] inicial para extração de termos foi composto pelas instruções apresentadas na Tabela 1.

Imagine que você é linguista, doutorado em regionalismo e cultura brasileira na escrita. Analise a letra de música abaixo, extraia expressões culturais e regionais que representam tristeza e depressão.

Entrada: letra de música

Saída: expressões regionais, expressões culturais Apresente em forma de tabela.

Entrada: inserir a letra de música

Tabela 1. Prompt 1 para extração de termos.

Entretanto, ajustes no prompt foram necessários pois, com os comandos passados no prompt 1 não era desconsiderado o contexto da música o que poderia influenciar a precisão dos termos extraídos. O prompt após os ajustes pode ser verificado na Tabela 2. Após ajustes no prompt para excluir influências contextuais das músicas, obteve-se uma redução de termos extraídos de 216 para 172, indicando que o contexto da música afetava a identificação dos termos.

Imagine que você é linguista, doutorado em regionalismo e cultura brasileira na escrita. Analise a letra de música abaixo, extraia expressões culturais e regionais que representam tristeza e depressão.

Não considere o contexto da música. Considere que isso é um trabalho científico e considere palavras de sentido de morte e suicídio.

Entrada: letra de música Saída: expressões regionais, expressões culturais.

Apresente em forma de tabela as expressões extraídas em uma única coluna chamada: Expressões Culturais e Regionais.

Caso a letra não apresente nenhuma expressão relacionada, escreva na coluna "sem expressão". Entrada: letra da música

Tabela 2. Prompt 2 para extração de termos.

- **Coleta de mensagens do Reddit:** Foram extraídas 10 mensagens da comunidade "Desabafos" no Reddit Brasil [Reddit 2024], um espaço público voltado ao compartilhamento de experiências pessoais. Essa amostra inicial serviu para testar a eficácia do léxico na identificação de indícios depressivos em texto.
- **Validação Inicial:** Os termos extraídos das letras de músicas, formando o léxico com termos culturais e regionais, foram utilizados para analisar mensagens coleadas da rede social Reddit. Para analisar as mensagens coletadas com ambos os léxicos, foi utilizado o prompt de comando no ChatGPT 3.5 apresentado na Tabela 3. Comparando com um léxico proposto por [Mendes 2024], observou-se que o léxico enriquecido culturalmente proposto nesse trabalho mostrou maior assertividade na detecção de sinais de depressão em texto quando comparado com a avaliação da especialista em saúde mental, tendo uma taxa de acertividade de 80% e o léxico proposto por [Mendes 2024] tendo uma taxa de 50%. Com esses indicativos, prosseguiu-se para o experimento final.

Utilize os termos acima que expressam tristeza e depressão, em seguida classifique se a mensagem é:
 Depressiva: sim ou não
 Se depressiva: qual o grau de depressão poderia ser atribuído baseado no PHQ-9
 Quais termos foram usados para identificar depressão na mensagem
 Mensagem:

Tabela 3. Prompt de identificação de mensagem depressiva segundo os termos do léxico em [Mendes 2024] e o léxico enriquecido culturalmente.

3.2. Experimento Final

Nesta seção, serão detalhados os passos realizados para a execução do experimento final do projeto, incluindo a incorporação de um maior volume de mensagens coletadas da rede social Reddit além do processo de extração e classificação das mensagens. O experimento final foi dividido nas seguintes etapas:

- **Construção do léxico:** Foram coletadas letras de músicas brasileiras que alcançaram o top 20 no Spotify, YouTube Music e TikTok nas duas primeiras semanas de outubro. Após remover duplicidades, 80 músicas foram selecionadas, sendo 78 disponíveis integralmente. O texto foi limpo e analisado [Spitzer et al. 1999] via API do GPT 3.4 para extrair termos relevantes conforme as nove categorias do PHQ-9¹, como em [Mendes 2024]. Os termos foram salvos em um arquivo .txt para posterior análise em Python, permitindo identificar expressões culturais e regionais ligadas às emoções.
- **Coleta das mensagens:** Foram coletadas 50 mensagens das comunidades brasileiras "Desabafos" (528 mil membros) e "Ansiedade e Depressão" (1,4 mil) no Reddit, plataforma escolhida por seu formato anônimo, que favorece a expressão emocional autêntica. A quantidade foi definida por uma especialista em saúde mental como adequada ao escopo do projeto. As mensagens (entre 2022 e 2024) foram avaliadas e anonimizadas para garantir a privacidade dos autores.
- **Uso das mensagens com GPT-3.4 e Flash 1.5:** As mensagens foram analisadas pelos LLMs GPT-3.4 e Flash 1.5, classificando grau de depressão (escala PHQ-9) e termos associados, tanto com o léxico original de [Mendes 2024] quanto com o léxico culturalmente adaptado aqui proposto. Para evitar perda de contexto com muitas mensagens, utilizou-se a técnica *Retriever-Augmented Generation* (RAG) com a biblioteca Chromadb. O uso de embeddings garantiu correlação precisa entre termos e mensagens. Com isso, os modelos mantiveram o contexto completo, assegurando consistência e integridade na classificação.

4. Resultados

A análise dos resultados foi conduzida em duas rodadas experimentais (R1 e R2):

¹As categorias são: falta de interesse, tristeza/humor deprimido, desordem de sono, falta de energia, desordem alimentar, baixa autoestima, problemas de concentração, hiperatividade/baixa atividade e pensamentos de suicídio.

4.1. R1 (Primeira Rodada)

Para a primeira rodada do experimento final foi utilizado o prompt de comando da Tabela 4 nos modelos GPT 3.4 e Flash 1.5 para identificação de traços de depressão nas mensagens coletadas.

Utilize o contexto para classificar a mensagem, não use nenhuma informação adicional, em seguida classifique se a mensagem é: Depressiva: sim ou não Calcule a pontuação segundo o PHQ-9 Se depressiva: qual o grau de depressão poderia ser atribuído baseado no PHQ-9 Quais termos foram usados para identificar depressão na mensagem.

Tabela 4. Prompt para classificar mensagens do R1.

Obteve-se os seguintes resultados para as mensagens analisadas se comparado com a análise da especialista em saúde mental:

- O léxico de [Mendes 2024] resultou em 64% de acertos para GPT 3.4 e 68% para Flash 1.5.
- O léxico culturalmente enriquecido aumentou a precisão para 78% (GPT 3.4) e 72% (Flash 1.5).
- No entanto, houve uma alta taxa de falsos positivos, comprometendo a especificidade da classificação.

Para resolver o problema de altas taxas de falsos positivos, foi necessário um ajuste no prompt de comando para que ele pudesse diferenciar mensagens depressivas de desabafos comuns.

4.2. R2 (Ajuste no Prompt)

Foram introduzidas instruções adicionais no prompt de comando visto na Tabela 4 para diferenciar mensagens depressivas de desabafos comuns.

O novo prompt de comando seguiu com as seguintes especificações como mostrado na Tabela 5.

Utilize o contexto para classificar a mensagem, não use nenhuma informação adicional. Verifique o relato da pessoa e tente associar se trata-se de algo possivelmente depressivo ou se é apenas um relato, desabafo, frustração, preocupação ou estresse passageiro sempre considere o contexto e não use nenhuma informação adicional. Em seguida classifique se a mensagem é: Depressiva: sim ou não Calcule a pontuação segundo o PHQ-9 Se depressiva: qual o grau de depressão poderia ser atribuído baseado no PHQ-9 Quais termos encontrados na mensagem foram usados para identificar depressão no contexto.

Tabela 5. Prompt para classificar mensagens do R2.

Com essas mudanças, a segunda fase dos experimentos (R2) apresentou uma melhoria na especificidade, reduzindo os falsos positivos e melhorando a distribuição entre verdadeiros positivos e negativos. As métricas de precisão e acurácia também melhoraram, tornando o modelo mais equilibrado e eficaz, passando de 64% para 88% de acurácia e 66.6% para 87% de precisão no GPT 3.4 e de 62% para 76% de acurácia e 63.8% para 74.2% de precisão para o Flash 1.5.

5. Discussão e Conclusão

Este estudo avaliou a eficácia de dois léxicos na identificação de mensagens depressivas: o léxico de [Mendes 2024] e um léxico adaptado com expressões culturais brasileiras.

Na fase R1, o léxico de [Mendes 2024] apresentou limitações com o prompt utilizado, com alta taxa de falsos positivos, principalmente no modelo Flash 1.5. Isso sugere que expressões regionais podem ser mal interpretadas por léxicos genéricos. O léxico adaptado reduziu esse problema, permitindo melhor distinção entre mensagens depressivas e não depressivas, especialmente no GPT 3.4, que passou de 32 para 39 acertos.

Na fase R2, ajustes no prompt aumentaram a especificidade dos modelos, reduzindo falsos positivos e melhorando o equilíbrio na classificação. A análise confirmou a importância de métricas como sensibilidade e especificidade para o refinamento dos modelos, evidenciando a necessidade de estratégias que minimizem classificações incorretas.

Comparando os modelos, o GPT 3.4 demonstrou maior sensibilidade às adaptações linguísticas, enquanto o Flash 1.5 manteve tendência a superestimar mensagens depressivas com o prompt usado. Além disso, o formato de saída do Flash 1.5 não seguiu o prompt, exigindo ajustes futuros.

Apesar dos avanços, a eficácia do léxico adaptado depende de atualização constante para refletir variações regionais e novas expressões. Estudos futuros podem envolver mais especialistas em saúde mental e explorar técnicas de engenharia de prompt para aprimorar a detecção de traços depressivos.

Os resultados reforçam a importância de ajustar os modelos com métricas adequadas e refinamentos no prompt. A construção contínua de léxicos adaptados e o uso de abordagens híbridas com aprendizado de máquina podem aperfeiçoar a análise de sentimentos, tornando-a mais sensível às particularidades linguísticas brasileiras.

6. Agradecimentos

Este trabalho foi parcialmente financiado pelo projeto “Pesquisa e Desenvolvimento de Algoritmos para Construção de Componentes Tecnológicos de Humanos Digitais” apoiado pelo Centro de Competência EMBRAPII em Tecnologias Imersivas com recursos financeiros oriundos do PPI IoT/Manufatura 4.0 / PPI HardwareBR do MCTI, através do Termo de Cooperação 057/2023, firmado com a EMBRAPII.

Referências

- Canaltech (2024). O que são prompts no chatgpt? Acessado em: 01 março de 2025.
- FIA (2024). O que é chatgpt? Acessado em: 01 março de 2025.
- Mendes, A.R. e Caseli, H. d. M. (2024). Detecção automática de postagens possivelmente depressivas em redes sociais. Acessado em: 01 maio 2024.
- Reddit (2024). Reddit. Acessado em: 01 fevereiro de 2025.
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., and the Patient Health Questionnaire Primary Care Study Group (1999). Validation and utility of a self-report version of prime-md: The phq primary care study. *Journal of the American Medical Association*, 282(18):1737–1744.
- Spotify (2025). Spotify - plataforma de streaming de música. Acesso em: 16 mar. 2025.