

# A system for monitoring depression symptoms from posts on social networks

Rodrigo Carvalho<sup>1</sup>, Matheus Yasuo Ribeiro Utino<sup>2</sup>,  
Paulo Mann<sup>3</sup>, Rodrigo Salvador Monteiro<sup>1</sup>

<sup>1</sup> Instituto de Computação, Universidade Federal Fluminense (UFF)

<sup>2</sup>Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (USP)

<sup>3</sup>Instituto de Computação, Universidade Federal do Rio de Janeiro (UFRJ)

carvalho\_rodrigo@id.uff.br, matheusutino@usp.br

paulomannjr@gmail.com, salvador@ic.uff.br

**Abstract.** *This work presents the development of a system for monitoring depressive symptoms based on social media posts, with a focus on the Reddit platform. By combining Natural Language Processing techniques and machine learning models, the posts were classified according to their potential indication of depression, complemented by lexical and thematic analyses. The developed web application enables statistical visualization and real-time monitoring, offering features aimed at supporting mental health professionals. The results show a significant correlation between the content of the analyzed communities and the presence of depressive signs, suggesting the system's potential as an auxiliary tool for early symptom screening. Ethical issues related to privacy and classification accuracy are also discussed.*

**Resumo.** *Este trabalho apresenta o desenvolvimento de um sistema para monitoramento de sintomas depressivos a partir de publicações em redes sociais, com foco na plataforma Reddit. Por meio da combinação de técnicas de Processamento de Linguagem Natural e modelos de aprendizado de máquina, as publicações foram classificadas quanto ao potencial indício de depressão, sendo complementadas por análises léxicas e temáticas. A aplicação web desenvolvida permite visualização estatística e monitoramento em tempo real, com funcionalidades voltadas ao apoio de profissionais de saúde mental. Os resultados obtidos evidenciam uma correlação significativa entre o conteúdo das comunidades analisadas e a presença de sinais depressivos, sugerindo o potencial do sistema como ferramenta auxiliar na triagem precoce de sintomas. Questões éticas relacionadas à privacidade e à acurácia das classificações também são discutidas.*

## 1. Introdução

O número de pessoas com depressão ultrapassa 280 milhões em todo o mundo [World Health Organization 2017]. Trata-se de um transtorno caracterizado por humor deprimido persistente ou pela perda de interesse e prazer em atividades anteriormente consideradas agradáveis, ao longo de um período prolongado [American Psychiatric Association et al. 2013]. Esses sintomas impactam diretamente

diversas esferas da vida, comprometendo os relacionamentos familiares, sociais e profissionais do indivíduo [World Health Organization 2023]. Diante de sua abrangência e profundidade, a depressão é considerada pela Organização Mundial da Saúde (OMS) a principal causa de incapacidade global, sendo também um grande fator de risco para o suicídio [World Health Organization 2017].

Dada a gravidade da depressão, é fundamental dispor de técnicas capazes de fornecer um rastreio preliminar, viabilizando o início oportuno de um tratamento adequado [Force 2023]. No entanto, diversos obstáculos dificultam esse processo, como a escassez de recursos financeiros, o estigma social relacionado aos transtornos mentais e a ausência — ou deficiência — na qualificação de profissionais da saúde [OPAS 2022]. Diante dessas barreiras, muitas pessoas afetadas acabam recorrendo a meios menos formais para expressar seus sentimentos e buscar algum tipo de apoio, sendo as redes sociais um espaço frequentemente utilizado para esse fim [Tadesse et al. 2019].

No estudo conduzido por [Seabrook et al. 2016], identificam e resumizam pesquisas relacionadas à influência das redes sociais na saúde mental dos internautas. No estudo referenciado, constatou-se que as interações realizadas nas redes sociais e o ambiente no qual o indivíduo interage estão correlacionadas tanto com a psicopatologia, quanto com o seu bem-estar. Outro estudo, centrado no fenômeno da depressão no Facebook, observou que a comparação social do próprio usuário em relação a outros usuários está fortemente relacionada à autodepreciação, que pode levar à depressão [Yoon et al. 2019].

Diante desse cenário, este trabalho propõe o desenvolvimento de um sistema para monitoramento de sintomas de depressão a partir de publicações em redes sociais, com o objetivo de contribuir para o rastreio preliminar de sintomas de depressão. O objetivo da ferramenta é ser utilizada por profissionais da saúde para complementar sua perspectiva clínica sobre o paciente, que deve estar ciente em disponibilizar o acesso, caso assim o deseje, ao médico ou psicólogo que o acompanha. Considerando a abundância de informações textuais e o relativo anonimato dos usuários, o Reddit foi adotado como fonte de dados. Trata-se de um agregador de conteúdo organizado em comunidades temáticas (*subreddits*), que facilita a segmentação por assunto. As publicações coletadas foram analisadas por um modelo de aprendizado de máquina com foco em Processamento de Linguagem Natural (PLN), combinando análise léxica e classificação de sentimentos.

Observamos, a partir das estatísticas providas pelo sistema, uma relação clara entre o teor da comunidade e as pesquisas literárias sobre a depressão, no que se refere a padrões de linguagens e comportamento de usuários com depressão em mídias sociais. Comunidades onde o assunto abordado é a própria depressão, têm como resultado 66% das publicações extraídas, classificadas como potencialmente depressivas e, em contrapartida, uma comunidade que aborda temas sobre economia, resulta em apenas 5%.

## **2. Trabalhos Relacionados**

O estudo de [Yazdavar et al. 2017] propõe uma abordagem estatística semi-supervisionada para rastrear sintomas depressivos em usuários do X (anteriormente conhecido como Twitter). A técnica se baseia nos critérios clínicos do questionário PHQ-9, e os autores coletaram mais de 23 milhões de tweets de 45.000 usuários autodeclarados como deprimidos. As contribuições incluem a construção de um léxico sintomático e o desenvolvimento de um modelo capaz de monitorar a progressão dos

sintomas ao longo do tempo. O modelo alcançou 68% de acurácia e 72% de precisão na detecção temporal de sintomas de depressão.

O trabalho de [Angskun et al. 2022] propõe um modelo de rastreamento de depressão em tempo real com base em dados de redes sociais, combinando características demográficas com análise de sentimentos extraída de tweets. Utilizando informações provenientes do questionário clínico PHQ-9, dados pessoais e interações no X, os autores construíram um modelo supervisionado avaliado com múltiplas técnicas de aprendizado de máquina, incluindo *Random Forest*, *SVM* e *Deep Learning*. O sistema é capaz de monitorar automaticamente o estado emocional dos usuários e emitir alertas sobre sinais de risco.

Em [Naseem et al. 2022], os autores introduzem uma classificação ordinal para mensurar o grau de severidade da depressão em publicações do Reddit, superando abordagens binárias. A categorização é feita em quatro níveis (mínimo, médio, moderado e severo) e utiliza um modelo hierárquico baseado em mecanismos de atenção. Foram analisadas 3.553 publicações rotuladas, reformuladas para refletir a gravidade da condição. O estudo demonstra que o uso de múltiplas publicações por usuário melhora a acurácia da predição do nível depressivo.

O trabalho de [Martínez-Castaño et al. 2020] apresenta uma plataforma escalável de análise em tempo real de sinais de depressão em mídias sociais, com foco em aplicações práticas para triagem e monitoramento em larga escala. A proposta envolve o desenvolvimento de uma biblioteca capaz de processar grandes volumes de dados e emitir alertas em tempo real com base em publicações extraídas do Reddit. Um protótipo funcional foi implementado como estudo de caso, com o objetivo de apoiar especialistas na identificação preliminar de transtornos mentais e nos tópicos mais recorrentes entre indivíduos com sintomas de depressão.

[Rosa et al. 2016] propõe um sistema de monitoramento baseado em análise de sentimentos, incorporando fatores de correção por idade e gênero, além de dados médicos autorizados. O modelo identifica indícios de sintomas de depressão em frases contendo construções como “Eu me sinto triste”, a partir de publicações no X. O sistema envia notificações a responsáveis autorizados sempre que há sinais de risco nas publicações. O estudo contou com 29 participantes durante 19 semanas e obteve um coeficiente de correlação de Pearson de 96,2 na associação entre os sentimentos expressos e os registros médicos.

De modo geral, os estudos analisados compartilham o objetivo de rastrear sintomas de depressão a partir de conteúdos publicados em redes sociais, com ênfase na modelagem computacional dos sintomas. Entretanto, poucos trabalhos avançam no desenvolvimento de sistemas interativos. Dentre eles, destacam-se os protótipos de [Martínez-Castaño et al. 2020] e [Rosa et al. 2016], embora com limitações em usabilidade e visualização. Diferentemente dessas propostas, nosso estudo se concentra na construção de uma plataforma web com interface dedicada, visualização estatística e recursos projetados para apoiar o trabalho de profissionais da saúde mental, como será detalhado nas seções seguintes.

### 3. Metodologia

A metodologia adotada neste trabalho envolve a coleta, processamento e análise de publicações em redes sociais com o objetivo de rastrear, preliminarmente, indivíduos com depressão. Para isso, foram empregadas técnicas de PLN e modelos de análise de sentimentos, integradas a uma aplicação web desenvolvida especificamente para monitoramento e visualização dos dados. A seguir, são descritas em detalhes as etapas do processo metodológico, incluindo a seleção e preparação do conjunto de dados, a arquitetura da aplicação desenvolvida e os métodos empregados para a classificação e análise das publicações.

#### 3.1. Conjunto de Dados

O conjunto de dados utilizado neste estudo foi obtido a partir da plataforma Reddit, uma rede social caracterizada pela organização de seu conteúdo em comunidades temáticas denominadas *subreddits*. A escolha do Reddit como fonte de dados se justifica por duas principais razões: (i) a estrutura temática de seus *subreddits*, que facilita a triagem e análise de conteúdo textual com base em tópicos específicos, e (ii) o acesso público aos dados por meio de uma API oficial, o que permite a coleta automatizada e transparente das publicações.

Para a composição do dataset, foram selecionados seis *subreddits* distintos no idioma Inglês americano. Três deles — *depression* (7.080 publicações), *Vent* (7.598 publicações) e *rant* (2.041 publicações) — abordam conteúdos predominantemente relacionados a desabafos, emoções negativas e saúde mental, sendo, portanto, mais propensos à manifestação de sintomas depressivos. Os demais — *ApplyingToCollege* (5.311 publicações), *movies* (3.495 publicações) e *economy* (1.399 publicações) — foram escolhidos como grupos de controle por tratarem de temas variados e não diretamente associados à depressão.

Essa seleção proporciona um conjunto de dados diversificado, contemplando diferentes níveis de expressão emocional e temas variados, o que é essencial para a construção de modelos robustos capazes de distinguir conteúdos relacionados à saúde mental de contextos cotidianos. A ampla quantidade de publicações coletadas contribui para a representatividade do dataset, garantindo uma base sólida para análises futuras.

#### 3.2. Aplicação Web

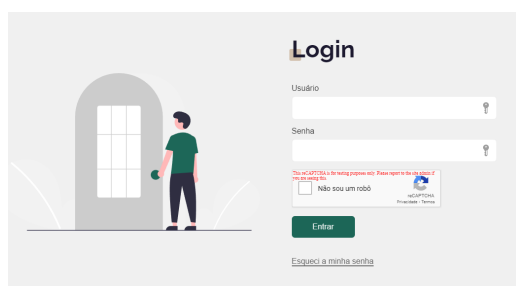
A aplicação web foi desenvolvida para viabilizar a coleta, análise e visualização de dados provenientes do Reddit, com foco na classificação de publicações quanto ao potencial indicativo de depressão. A arquitetura adotada combina uma API RESTful, implementada com o microframework Flask, e uma interface gráfica reativa, desenvolvida em React.js.

A escolha do Flask justifica-se por sua leveza, flexibilidade e compatibilidade com bibliotecas Python especializadas em processamento de linguagem natural e aprendizado de máquina, facilitando a integração com os modelos utilizados na análise textual. O React.js foi selecionado pela sua eficiência na construção de interfaces dinâmicas e responsivas, proporcionando uma experiência de usuário fluida. A comunicação entre frontend e backend ocorre via requisições REST, assegurando a modularidade e escalabilidade do sistema. A persistência dos dados é realizada em um banco relacional PostgreSQL, escolhido por sua robustez, suporte a transações e facilidade de integração com a stack tecnológica.

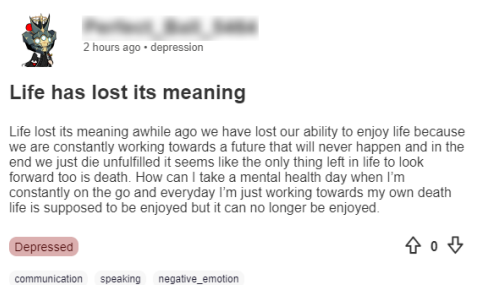
Para garantir segurança e controle de acesso, implementou-se autenticação baseada em papéis, segmentando permissões conforme o perfil do usuário (administrador, profissional da saúde ou perfil monitorado). As senhas são criptografadas com SHA-256, e a proteção contra ataques automatizados é realizada por meio do reCAPTCHA.

A interface web oferece funcionalidades como visualização de linhas do tempo com publicações extraídas dos *subreddits* monitorados, painéis estatísticos, mecanismos de busca e ferramentas de monitoramento de perfis. A atualização das publicações ocorre automaticamente a cada dez minutos, assegurando a atualidade das informações.

Adicionalmente, o sistema permite a avaliação das classificações realizadas pelo modelo, viabilizando a coleta de *feedback* para o aprimoramento contínuo da análise automatizada. Essa estrutura modular, baseada em tecnologias consolidadas, garante uma aplicação escalável, segura e de fácil manutenção, adequada aos objetivos do projeto. Para mais detalhes de implementação e sobre a interface visual acesse o repositório<sup>1</sup>.



(a) Tela de login do sistema.



(b) Postagem classificada do Reddit.

**Figura 1. Exemplo de duas telas do sistema web: interface de login e visualização de postagens classificadas.**

Além da implementação técnica, destaca-se que a aplicação foi concebida para ser utilizada em cenários reais de monitoramento e apoio à saúde mental. A estrutura de coleta automática e atualização periódica das publicações simula o funcionamento de um feed de notícias, no qual novas informações chegam continuamente ao sistema, permitindo que profissionais da saúde possam acompanhar em tempo quase real as manifestações textuais de indivíduos sob monitoramento.

Nesse contexto, os experimentos realizados com o sistema visam não apenas validar a acurácia da classificação automatizada, mas também demonstrar a sua aplicabilidade prática como ferramenta de triagem e suporte clínico. A proposta é que, ao receber notificações ou visualizar atualizações no painel do sistema, o profissional da saúde possa avaliar rapidamente o conteúdo das postagens recentes e, a partir disso, tomar decisões informadas sobre a necessidade de intervenção ou encaminhamento do paciente para um atendimento especializado.

Adicionalmente, ao permitir a avaliação manual das classificações geradas pelo modelo, o sistema também se configura como um recurso de apoio à decisão, promovendo um ciclo virtuoso de aprendizado: o profissional pode confirmar ou corrigir a análise automatizada, contribuindo assim para o aprimoramento contínuo do modelo de detecção. Dessa forma, o sistema se aproxima de um ambiente realista de uso, onde tecnologia e

<sup>1</sup><https://osf.io/rkxne/>

expertise humana se complementam no enfrentamento de um problema de alta relevância social e clínica.

A importância dessa abordagem reside na possibilidade de antecipar sinais de sofrimento psíquico, oferecendo aos profissionais de saúde mental uma ferramenta tecnológica que amplia sua capacidade de detecção precoce de potenciais quadros depressivos. Em muitos contextos, especialmente na atenção primária e em serviços sobrecarregados, a identificação de mudanças sutis no comportamento dos pacientes pode ser dificultada pela limitação de recursos e tempo. O sistema proposto atua, portanto, como um meio de monitoramento contínuo, permitindo que os profissionais acompanhem a evolução do estado emocional dos indivíduos de forma sistemática e eficiente, o que pode favorecer intervenções mais rápidas e adequadas, reduzindo o risco de agravamento dos quadros depressivos e, conseqüentemente, de eventos críticos como ideação suicida.

### **3.3. Classificação das Publicações**

Para a classificação das publicações, utilizou-se o modelo de análise de sentimentos *Valence Aware Dictionary and sEntiment Reasoner* (VADER) [Hutto and Gilbert 2014], que destaca-se por incorporar a habilidade humana em seu processo de desenvolvimento e por sua elevada eficácia na análise de textos oriundos de mídias sociais, dispensando a necessidade de dados adicionais para treinamento supervisionado. Essa característica o torna especialmente adequado para aplicações em ambientes com fluxo contínuo de dados, já que é capaz de operar com rapidez suficiente para processamento online em streaming.

O modelo VADER realiza a análise da polaridade do sentimento, fornecendo uma pontuação em uma escala que varia de -4 (extremamente negativo) a +4 (extremamente positivo), incluindo o valor zero para sentimentos neutros. Neste estudo, as publicações foram consideradas potencialmente indicativas de sintomas depressivos quando classificadas com polaridade negativa na faixa de -1 a -4. É importante destacar que a metodologia adotada não é restrita ao uso do VADER, sendo passível de incorporação de outros modelos de classificação, conforme a necessidade.

Além disso, para a categorização léxica e análise temática das publicações, utilizou-se a ferramenta Empath [Fast et al. 2016]. O Empath consiste em um léxico dinâmico extraído de textos contemporâneos da web, similar ao reconhecido *Linguistic Inquiry and Word Count* (LIWC) [Pennebaker et al. 2001], mas com a vantagem de possuir categorias transparentes, facilmente extensíveis e de rápida aplicação. O Empath permite a geração de categorias lexicais que refletem a conotação do texto, por meio de valores de ocorrência que variam de 1 a 10, possibilitando uma análise complementar dos conteúdos publicados.

## **4. Resultados**

A seção será organizada em análise quantitativa e análise qualitativa, explorando os dados diretos sobre as classificações e analisando os padrões de comportamento e linguagem dos usuários. Em uma análise quantitativa dos dados recuperados, obtemos resultados variados de acordo com a comunidade analisada, demonstrando uma consonância com a temática abordada.

#### 4.1. Análise Quantitativa

Os dados quantitativos consolidados encontram-se na Tabela 1. A proporção de publicações classificadas como potencialmente depressivas nos *subreddits depression*, *Vent* e *rant*, em relação ao total de publicações extraídas de cada comunidade, corresponde, respectivamente, a 65.92%, 59.21% e 62.18%.

Em termos absolutos, no *subreddit depression*, das 7.080 publicações analisadas, 4.667 foram classificadas pelo modelo como potencialmente depressivas, enquanto as 2.413 restantes foram categorizadas como não depressivas. Para o *subreddit Vent*, entre as 7.598 publicações coletadas, 4.498 foram identificadas como depressivas e 3.100 como não depressivas. Já na comunidade *rant*, com um total de 2.041 publicações, 1.269 foram rotuladas como depressivas e 772 como não depressivas.

Com o objetivo de avaliar o comportamento do modelo em contextos temáticos distintos, foram analisadas também comunidades cujo foco não está diretamente associado a questões emocionais ou de saúde mental. No *subreddit ApplyingToCollege*, das 5.311 publicações extraídas, 1.278 (24.06%) foram classificadas como potencialmente depressivas. Na comunidade *movies*, foram coletadas 3.495 publicações, das quais 608 (17.40%) foram consideradas depressivas. Por fim, no *subreddit economy*, apenas 72 publicações de um total de 1.399 (5.15%) foram classificadas como potencialmente depressivas.

Os resultados obtidos corroboram a expectativa de que *subreddits* com temáticas mais sensíveis à saúde mental apresentem uma incidência significativamente maior de publicações potencialmente depressivas. Esse padrão pode ser, em parte, atribuído à natureza do Reddit como plataforma de discussão contínua, o que favorece o aprofundamento de tópicos delicados. Apesar disso, estudos indicam que emoções negativas tendem a se espalhar com mais intensidade e rapidez em ambientes digitais, sobretudo por meio de laços sociais fracos — ou seja, conexões entre usuários com pouca ou nenhuma familiaridade prévia. Esse tipo de vínculo, comum em redes sociais abertas, facilita a circulação de conteúdos emocionalmente intensos entre diferentes grupos, ampliando o alcance de sentimentos como tristeza, raiva ou desesperança. Além disso, usuários mais engajados tendem a manifestar uma intensificação gradual na negatividade de suas publicações, o que sugere um possível efeito cumulativo associado à exposição contínua a esse tipo de conteúdo [Del Vicario et al. 2016, Fan et al. 2016]. Nesse sentido, ainda que comunidades voltadas à saúde mental no Reddit sejam frequentemente estruturadas em torno do suporte mútuo, a própria lógica algorítmica e os padrões de engajamento dessas plataformas podem, paradoxalmente, contribuir para a manutenção ou o agravamento de estados emocionais negativos em indivíduos mais vulneráveis.

#### 4.2. Análise Qualitativa

A análise qualitativa dos dados extraídos permite uma compreensão mais profunda sobre os padrões de comportamento e o uso da linguagem por parte de usuários potencialmente depressivos em mídias sociais. A Tabela 2 apresenta as cinco categorias mais recorrentes em quatro comunidades distintas: duas associadas a alta incidência de publicações depressivas (*depression* e *Vent*) e duas com as menores taxas (*economy* e *movies*). Essa comparação reforça a hipótese de que o conteúdo semântico das publicações está diretamente relacionado à natureza temática da comunidade.

**Tabela 1. Números de publicações por comunidade.**

<i>Subreddit</i>	Total	Depressivas	Não depressivas	% Depressivas
Depression	7.080	4.667	2.413	65.92%
Vent	7.598	4.498	3.100	59.21%
Rant	2.041	1.269	772	62.18%
ApplyingToCollege	5.311	1.278	4.033	24.06%
Movies	3.495	608	2.887	17.40%
Economy	1.399	72	1.327	5.15%

Nas comunidades com maior incidência de publicações potencialmente depressivas, observa-se uma clara predominância de categorias léxicas que refletem sintomas e experiências comuns na depressão, indicando uma repetição significativa de temas associados ao sofrimento emocional. No *subreddit depression*, as categorias mais frequentes — “emoção negativa”, “sofrimento”, “tristeza”, “violência” e “vergonha” — refletem diretamente estados emocionais característicos da condição depressiva, como angústia, desesperança e sentimentos de autoavaliação negativa [National Institute of Mental Health (NIMH) 2024]. De forma semelhante, a comunidade *vent* também exibe a repetição de categorias como “emoção negativa”, “sofrimento” e “violência”, mas com também com o surgimento de “ódio” e “amigos”, o que sugere uma complexidade emocional maior, envolvendo tanto a expressão de emoções negativas intensas quanto a dinâmica social entre os usuários [National Institute of Mental Health (NIMH) 2024]. A presença recorrente de termos ligados à violência em ambas as comunidades indica a manifestação de sentimentos profundos de conflito interno e possíveis reações agressivas, que podem ser sintomas associados a quadros de sofrimento psicológico intenso [Fazel et al. 2015]. Essa repetição de categorias temáticas entre diferentes *subreddits* reforça a existência de padrões consistentes de expressão emocional e sintomatologia depressiva, evidenciando como esses espaços digitais funcionam como canais onde indivíduos vulneráveis externalizam experiências subjetivas similares.

Em contraste, as comunidades com menor incidência de publicações potencialmente depressivas revelam campos semânticos mais ligados a temas objetivos ou a interações sociais positivas. No *subreddit economy*, por exemplo, as categorias mais frequentes — “economia”, “dinheiro”, “negócios”, “bancário” e “pagamento” — refletem um foco em discussões práticas e informativas relacionadas ao universo financeiro, com pouca presença de carga emocional negativa. Essa objetividade sugere que os usuários dessa comunidade utilizam o espaço principalmente para troca de informações e debates técnicos, distanciando-se das expressões subjetivas de sofrimento emocional observadas em *subreddits* voltados à saúde mental. Já o *subreddit movies* apresenta uma diversidade maior nas categorias semânticas, incluindo “felicidade”, “emoção negativa”, “amigos”, “emoções positivas” e “jogar”. Essa variedade pode ser atribuída à natureza plural dos filmes, que abordam uma ampla gama de gêneros, temas e emoções, possibilitando tanto expressões de alegria e diversão quanto manifestações de sentimentos negativos, como tristeza ou frustração. Assim, o ambiente se configura como um espaço onde predominam interações sociais mais leves e recreativas, favorecendo a troca de experiências e



opiniões que envolvem diferentes tipos de conteúdo emocional, em contraste com a homogeneidade emocional observada em *subreddits* dedicados a temas depressivos.

Esses resultados evidenciam que o conteúdo linguístico das publicações está fortemente condicionado pelo tema central de cada comunidade, refletindo padrões emocionais e comportamentais distintos. Enquanto *subreddits* associados a altos índices de publicações potencialmente depressivas apresentam uma linguagem marcada pela repetição de sentimentos negativos e sintomas relacionados ao sofrimento mental, comunidades com foco em temas mais objetivos ou de entretenimento exibem uma diversidade temática e emocional maior, com predominância de interações menos carregadas emocionalmente. Tal contraste reforça a importância de compreender o contexto e a natureza dos ambientes digitais ao analisar manifestações de saúde mental online, destacando como diferentes espaços podem tanto refletir quanto influenciar os estados emocionais dos usuários.

**Tabela 2. Categorias mais frequentes entre publicações por comunidade.**

<i>Subreddit</i>	1ª Categoria	2ª Categoria	3ª Categoria	4ª Categoria	5ª Categoria
<i>Depression</i>	negative-emotion	pain	sadness	violence	shame
<i>Vent</i>	negative-emotion	pain	hate	friends	violence
<i>Economy</i>	economics	money	business	money	payment
<i>Movies</i>	fun	negative-emotion	friends	positive-emotion	play

A Tabela 3 apresenta a distribuição das publicações potencialmente depressivas ao longo dos diferentes períodos do dia, segmentadas por comunidade analisada. Para fins de categorização temporal, considerou-se como madrugada o intervalo entre 00h00 e 05h59; manhã, entre 06h00 e 11h59; tarde, entre 12h00 e 17h59; e noite, entre 18h00 e 23h59. Observa-se que os padrões temporais de atividade revelam aspectos relevantes do comportamento dos usuários. Especificamente, publicações classificadas como depressivas tendem a se concentrar nos períodos da noite e da madrugada. Tal padrão está em consonância com achados anteriores da literatura, como o estudo de [Vedula and Parthasarathy 2017], que identificou maior atividade noturna entre usuários com indícios de sintomas depressivos — fenômeno possivelmente associado a distúrbios do sono, comumente relacionados à depressão.

**Tabela 3. Quantidade de publicações depressivas de acordo com o período do dia.**

<i>Subreddit</i>	Madrugada	Manhã	Tarde	Noite
<i>Depression</i>	1.452	810	1.001	<b>1.555</b>
<i>Vent</i>	<b>1.504</b>	846	937	1.399
<i>Rant</i>	379	277	288	<b>420</b>
<i>ApplyingToCollege</i>	<b>445</b>	188	281	408
<i>Movies</i>	186	101	139	<b>209</b>
<i>Economy</i>	19	16	<b>35</b>	23

## 5. Implicações Éticas

A aplicação de métodos de aprendizado de máquina em contextos sensíveis, como a saúde mental, exige rigor técnico e um forte compromisso ético em todas as etapas do ciclo de vida do modelo. Questões como mitigação de vieses algorítmicos e responsabilidade clínica emergem como desafios centrais, requerendo a observância de princípios como transparência, justiça e precaução [Morley et al. 2020].

Modelos de aprendizado de máquina aprendem padrões a partir dos dados de treinamento e, por isso, podem refletir e até amplificar tendências ou vieses presentes nesses dados [Mehrabi et al. 2021]. No contexto da saúde mental, essa questão torna-se especialmente crítica, pois dados enviesados podem levar a inferências distorcidas ou imprecisas, potencialmente reforçando disparidades em diagnósticos, recomendações de tratamento ou acesso aos cuidados. Isso é particularmente preocupante para populações sub-representadas ou marginalizadas, cujas experiências podem não estar adequadamente capturadas nos dados, resultando em modelos que falham em generalizar corretamente ou que perpetuam desigualdades sistêmicas.

Embora sistemas automatizados possam desempenhar um papel relevante de apoio na triagem inicial, sua aplicação prática requer supervisão contínua por profissionais qualificados [Wang et al. 2025]. Limitações técnicas — como a dificuldade dos modelos baseados em texto em captar nuances contextuais (por exemplo, ironia ou sarcasmo) e fatores extralinguísticos (por exemplo, histórico médico) — tornam indispensável a mediação humana. Um falso positivo poderia acarretar estigmatização indevida, enquanto um falso negativo poderia atrasar cuidados essenciais. Por isso, propõe-se que as classificações produzidas pelo sistema sirvam apenas como alertas preliminares, subordinados às avaliações clínicas realizadas por profissionais certificados [Ghassemi et al. 2020]. Essa abordagem híbrida preserva a autonomia do paciente, assegura a conformidade com princípios bioéticos — como o não malefício — e mantém a responsabilidade legal dentro do escopo dos provedores de saúde licenciados.

## 6. Conclusão

Este trabalho propôs e implementou uma plataforma para monitoramento de sintomas depressivos com base em conteúdos textuais oriundos do Reddit. Os resultados demonstraram a eficácia do sistema na identificação de padrões linguísticos associados à depressão, tanto em análises quantitativas quanto qualitativas. A arquitetura modular da aplicação e o uso de ferramentas amplamente adotadas garantem sua escalabilidade e adaptabilidade a diferentes contextos.

Como trabalhos futuros, propõe-se a incorporação de modelos mais sofisticados de análise semântica, como transformers pré-treinados em dados clínicos, para melhorar a acurácia das classificações. Além disso, será considerada a expansão para outras redes sociais, como Instagram ou TikTok, permitindo uma análise multimodal que integre textos, imagens e vídeos. Também pretende-se validar o sistema em parceria com profissionais da saúde mental, a fim de verificar sua aplicabilidade clínica e promover melhorias guiadas por feedback especializado. Finalmente, é essencial aprofundar os mecanismos de explicabilidade do modelo e mitigar possíveis vieses, promovendo uma abordagem ética e transparente.

## Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Apoio adicional foi fornecido pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), sob o número de concessão 130041/2025-4.

## Referências

- American Psychiatric Association, A., Association, A. P., et al. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*, volume 10. Washington, DC: American psychiatric association.
- Angskun, J., Tipprasert, S., and Angskun, T. (2022). Big data analytics on social networks for real-time depression detection. *Journal of Big Data*, 9(1):69.
- Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., Stanley, H. E., and Quattrociocchi, W. (2016). Echo chambers: Emotional contagion and group polarization on facebook. *Scientific Reports*, 6:37825.
- Fan, R., Xu, K., and Zhao, J. (2016). Higher contagion and weaker ties mean anger spreads faster than joy in social media. *arXiv preprint arXiv:1608.03656*.
- Fast, E., Chen, B., and Bernstein, M. S. (2016). Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 4647–4657.
- Fazel, S., Wolf, A., Chang, Z., Larsson, H., Goodwin, G. M., and Lichtenstein, P. (2015). Depression and violence: a swedish population study. *The Lancet Psychiatry*, 2(3):224–232.
- Force, U. P. S. T. (2023). Screening for depression and suicide risk in adults: Us preventive services task force recommendation statement. *JAMA*, 329(23):2057–2067.
- Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., Chen, I. Y., and Ranganath, R. (2020). A review of challenges and opportunities in machine learning for health. *AMIA Summits on Translational Science Proceedings*, 2020:191.
- Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):216–225.
- Martínez-Castaño, R., Pichel, J. C., and Losada, D. E. (2020). A big data platform for real time analysis of signs of depression in social media. *International Journal of Environmental Research and Public Health*, 17(13).
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Comput. Surv.*, 54(6):1–35.
- Morley, J., Machado, C. C., Burr, C., Cows, J., Joshi, I., Taddeo, M., and Floridi, L. (2020). The ethics of ai in health care: a mapping review. *Soc Sci Med*, 260:113172.
- Naseem, U., Dunn, A. G., Kim, J., and Khushi, M. (2022). Early identification of depression severity levels on reddit using ordinal classification. In *Proceedings of the ACM Web Conference 2022, WWW '22*, page 2563–2572, New York, NY, USA. Association for Computing Machinery.

- National Institute of Mental Health (NIMH) (2024). Depression. <https://www.nimh.nih.gov/health/publications/depression>. Accessed: 2025-06-04.
- OPAS (2022). Depressão - opas/oms — organização pan-americana da saúde.
- Pennebaker, J., Francis, M., and Booth, R. (2001). Linguistic inquiry and word count (liwc): Liwc2001. 71.
- Rosa, R. L., Rodríguez, D. Z., Schwartz, G. M., de Campos Ribeiro, I., and Bressan, G. (2016). Monitoring system for potential users with depression using sentiment analysis. In *2016 IEEE International Conference on Consumer Electronics (ICCE)*, pages 381–382.
- Seabrook, E. M., Kern, M. L., and Rickard, N. S. (2016). Social networking sites, depression, and anxiety: a systematic review. *JMIR mental health*, 3(4):e5842.
- Tadesse, M. M., Lin, H., Xu, B., and Yang, L. (2019). Detection of depression-related posts in reddit social media forum. *IEEE Access*, 7:44883–44893.
- Vedula, N. and Parthasarathy, S. (2017). Emotional and linguistic cues of depression from social media. In *Proceedings of the 2017 International Conference on Digital Health, DH '17*, page 127–136, New York, NY, USA. Association for Computing Machinery.
- Wang, Y., Hu, W., Zhou, K., et al. (2025). What is the role of human decisions in a world of artificial intelligence: an economic evaluation of human-ai collaboration in diabetic retinopathy screening. *ArXiv*, abs/2503.20160.
- World Health Organization (2017). Depression and other common mental disorders: Global health estimates. Technical report, World Health Organization, Geneva. Licence: CC BY-NC-SA 3.0 IGO.
- World Health Organization (2023). Depression. <https://www.who.int/news-room/fact-sheets/detail/depression>. Accessed: 2025-05-21.
- Yazdavar, A. H., Al-Olimat, H. S., Ebrahimi, M., Bajaj, G., Banerjee, T., Thirunarayan, K., Pathak, J., and Sheth, A. (2017). Semi-supervised approach to monitoring clinical depressive symptoms in social media. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17*, page 1191–1198, New York, NY, USA. Association for Computing Machinery.
- Yoon, S., Kleinman, M., Mertz, J., and Brannick, M. (2019). Is social network site usage related to depression? a meta-analysis of facebook–depression relations. *Journal of Affective Disorders*, 248:65–72.