

Identificando sintomas depressivos: um estudo de caso no Youtube

Rodolpho da Silva Nascimento¹, Flavio Carvalho¹, Gustavo Paiva Guedes¹

¹CEFET/RJ - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca
Av. Maracanã, 229 - Rio de Janeiro - RJ - Brasil.

{rodolpho.nascimento, flavio.carvalho}@eic.cefet-rj.br,
gustavo.guedes@cefet-rj.br

Abstract. *Depression is an illness that has taken worrying proportions all over the world. According to the World Health Organization, more than 300 million people suffer from this disease. With popularization and facilities became from the Internet, it's common to find depressive peoples that in attempts to relieve their symptoms, end up exposing their feelings in social networks before reaching the extreme of suicide. This paper proposes the use of a lexicon to improve the identification of depressive symptoms in a Brazilian social network. Preliminary experiments indicated that using the lexicon leads to an improvement in the results in the classification tasks for the detection of depression symptoms on texts.*

Resumo. *A depressão é uma doença que vem tomando proporções preocupantes em todo o mundo. Segundo a Organização Mundial de Saúde, mais de 300 milhões de pessoas sofrem deste mal. Com a popularização e facilidades de acesso à Internet nos dias atuais, é comum encontrar indivíduos depressivos que, na tentativa de aliviar seus sintomas, acabam expondo seus sentimentos em redes sociais antes de chegar ao extremo do suicídio. Este trabalho propõe o uso de um léxico para aprimorar a identificação de sintomas depressivos em redes sociais brasileiras. Experimentos preliminares indicaram que a utilização do léxico melhoram os resultados nas tarefas classificação para detecção de sintomas de depressão em textos.*

1. Introdução

No Brasil, o uso da Internet tem se tornado cada vez mais frequente na população. Uma recente pesquisa realizada pelo IBGE indica que a utilização da Internet nos domicílios vem crescendo rapidamente, descrevendo que, em 2016, a Internet era utilizada em 69,3% dos domicílios permanentes do País e, no ano seguinte, este percentual aumentou para 74,9% [PNAD, 2017]. Essa pesquisa também destaca que 98,7% das pessoas utilizam telefones celulares para acessar a Internet.

Ainda segundo o IBGE, das pessoas que utilizam Internet (126 milhões), 95.5% afirmam enviar ou receber mensagens de textos que não sejam *e-mails* (e.g., redes sociais) [IBGE, 2017]. Nesse cenário e com a grande adesão de usuários conectados a Internet, além das facilidades de acesso, se torna cômoda e ágil a contribuição para a exposição de conteúdos em redes sociais.

As redes sociais possuem grande representação na comunicação social nos dias de hoje, em que o mundo virtual tem se tornado cada vez mais uma extensão da vida cotidiana. Os indivíduos constantemente desenvolvem o hábito de expor seus pensamentos e seus sentimentos nessas redes. Por exemplo, um indivíduo depressivo busca de muitas formas aliviar seus sentimentos e, na busca por um conforto, acaba expondo seus sentimentos em tais plataformas [Silveira et al., 2018].

Não raro, casos de indivíduos expondo sentimentos depressivos em redes sociais têm-se tornado comum e alguns, infelizmente, terminaram de forma trágica. Podemos citar o caso de um jovem de 17 anos que, em outubro de 2018, cometeu suicídio no interior da loja onde trabalhava, no entanto, antes disso, deixou uma carta de despedida no Facebook¹. Tal caso tornou-se mais trágico pois dias depois, seu amigo de 18 anos cometeu suicídio após visualizar a mensagem no Facebook. Ele também deixou uma carta de despedida no Facebook lamentando a perda do amigo².

A depressão é uma doença caracterizada por sentimentos negativos de um indivíduo, como tristeza constante, perda de interesse em suas atividades cotidianas, sentimentos de inutilidade, insônia e recorrentes pensamentos suicidas [DSM-V, 2013]. Nas últimas décadas, essa doença tem tomado proporções preocupantes, conforme se observa em dados fornecidos pela Organização Mundial de Saúde (OMS). De acordo com a OMS, somente entre os anos 2005 e 2015, houve um crescimento de mais de 18% de indivíduos diagnosticados com depressão, e, atualmente, mais de 300 milhões sofrem desse mal [WHO, 2017]. Tais números serviram de impulso para a OMS fomentar o lançamento de campanhas relacionadas ao diagnóstico e tratamento da depressão por todo mundo [WHO, 2015].

No Brasil, estima-se que 7,6% das pessoas com 18 anos ou mais já receberam diagnóstico de depressão por algum profissional de saúde mental, representando 11,2 milhões de pessoas, com maior prevalência na área urbana (8,0%) do que rural (5,6%) [IBGE, 2013]. As Regiões Sul e Sudeste apresentam os maiores percentuais de pessoas com depressão diagnosticada, acima do percentual nacional, 12,6% e 8,4%, respectivamente [IBGE, 2013].

Em razão do exposto nos parágrafos precedentes, nota-se a importância do desenvolvimento de pesquisas científicas que encontrem soluções para identificar eficazmente sinais de depressão em textos em português do Brasil, atuando como suporte para grupos de apoio a depressão. Nesse cenário, o presente trabalho contribui com a utilização de um léxico para aprimorar os resultados dos algoritmos da área de Aprendizagem de Máquina na tarefa de identificar sintomas depressivos em *posts* de redes sociais. Além disso, contribui com a construção e anotação de um conjunto de dados com textos provenientes da rede social Youtube, o que pode servir como base para trabalhos que envolvam a detecção de sintomas depressivos em textos em português do Brasil.

As demais seções estão dispostas da seguinte maneira: na Seção 2, são levantados trabalhos semelhantes e relacionados ao tema; na Seção 3, são apresentados os passos utilizados para a criação do conjunto de dados proposto nesse estudo; na Seção 4, é apre-

¹<https://noticiasguariba.com.br/blog/conteudo-geral/noticias-gerais/adolescente-de-17-anos-posta-mensagem-no-facebook-e-em-seguida-comete-suicidio>

²<http://www.jornalrazaa.com/noticias/policia/jovem-comete-suicidio-apos-morte-de-amigo-1.2105658>

sentada a metodologia para análise do conjunto de dados; na Seção 5, são discutidos os resultados encontrados; por último, a Seção 6 apresenta as conclusões, limitações e discussão sobre trabalhos futuros.

2. Trabalhos relacionados

Chen et al. [2018] realizaram a coleta de dados do Twitter³ durante quatro meses, utilizando como *string* de busca o conteúdo “*I was/have been diagnosed with depression*” visando encontrar relatos de depressão. Em seguida, para cada usuário identificado no grupo selecionado, foram coletadas todas as postagens anteriores, limitando-se ao período de um ano. Ao final, foram selecionados 600 usuários únicos com média de 2.000 *posts* para cada. As características emocionais foram extraídas dos textos e para a classificação binária (*índice de depressão = S/N*) foram utilizados os classificadores SVM e Random Forest. Em seguida foram utilizadas séries temporais para avaliações emocionais.

Nguyen et al. [2014] coletaram 38.401 *posts* de comunidades com interesse em características depressivas da rede social LiveJournal⁴, utilizando na pesquisa palavras-chave como *depression, bipolar, self-harm, attachment/separation e suicide*. De forma complementar, selecionaram 229.563 *posts* dos mais variados assuntos. Para extrair as características afetivas, foi utilizado o léxico ANEW [Bradley and Lang, 1999], que possui 1.034 palavras avaliadas em termos de valência e excitação. Para a extração de características linguísticas, foi utilizado o LIWC [Pennebaker, 2011]. Modelos de regressão logística foram utilizados para a seleção de características dos *posts*. Os resultados indicaram alto grau de relacionamento entre *posts* depressivos e *posts* de pensamentos suicidas. Os autores também destacam o grande potencial da aplicabilidade de *machine learning* em pesquisas psiquiátricas.

Tsugawa et al. [2015] elaboraram um questionário *online* (CES-D, Radloff [1977]) para avaliar o grau de depressão em um grupo de usuários, sendo disponibilizado através de um link em um site próprio. O link foi publicado na rede social Twitter e que os usuários que apresentaram interesse em preencher o questionário foram previamente apresentados aos objetivos da pesquisa, além de concordarem explicitamente em conceder acesso aos seus *posts* publicados, o que gerou, em média, 3.200 *posts* por usuário. Um léxico (elaborado pelos autores) contendo 862 palavras negativas e 760 positivas foi utilizado no classificador. Segundo os autores, a ausência de um léxico padrão na língua japonesa (*e.g.* ANEW [Bradley and Cuthbert, 1999]) pode ter afetado o desempenho do modelo apresentado.

Dados coletados das redes sociais LiveJournal⁵, Facebook⁶ e Twitter foram utilizados no trabalho de Aldarwish and Ahmad [2017]. Para criar um modelo preditivo, utilizaram a ferramenta RapidMiner⁷ com os classificadores Naive Bayes e SVM em *posts* depressivos (2.073 entradas) e não-depressivos (4.700) treinados manualmente, conforme sintomas definidos pela *American Psychiatric Association Diagnostic and Statistical Manual* (DSM-IV [2011]). Com uma acurácia de 57%, os autores justificam como razões

³<https://www.twitter.com/>

⁴<https://www.livejournal.com/>

⁵<https://www.livejournal.com/>

⁶<https://www.facebook.com/>

⁷<https://rapidminer.com/>

para o baixo desempenho do modelo a dificuldade de diferenciar o mesmo usuário ativo no Twitter e Facebook ao mesmo tempo, além da dificuldade da linguagem árabe, em que uma palavra pode ter diferentes significados.

Nenhum dos trabalhos supracitados utilizou um léxico elaborado exclusivamente com termos de cunho depressivo, buscando aprimorar o resultado dos experimentos. Também utilizaram a rede social Youtube como base de dado para coletar relatos com indícios depressivos e extrair inferências com maior acurácia.

3. Criação do conjunto de dados

Essa seção descreve o conjunto de dados criado a partir de publicações no Youtube, com o objetivo de servir como base para trabalhos em português do Brasil que envolvam a detecção de sintomas relacionados à depressão em textos. A Seção 3.1 fundamenta a escolha da rede social utilizada para a criação desse conjunto de dados. A Seção 3.2 descreve o processamento efetuado para a aquisição dos dados utilizados para formar esse conjunto de dados. Por fim, a Seção 3.3 descreve os passos utilizados para a rotulação manual dos textos.

3.1. Escolha da rede social

O trabalho proposto neste artigo consiste em utilizar a rede social Youtube⁸ para a extração de comentários sobre vídeos relacionados à depressão e classificar os comentários como *indício de depressão* = S/N . A motivação em usar essa rede social deu-se na observância do amplo poder de alcance e pela baixa explorabilidade da plataforma em pesquisas do tipo. Inúmeros canais no Youtube registram milhares de inscritos, e o conteúdo publicado pelos inscritos podem, em poucas horas, registrar grande quantidade de visualizações⁹, denotando assim o grande poder de alcance na comunicação com seus inscritos.

Podemos citar o canal da famosa *youtuber* Renata Castilho¹⁰, que possui 385.000 usuários inscritos. Somente em um vídeo publicado com o título “SUICÍDIO e Depressão!”, há o registro de 1.343.042 de visualizações e de 7.532 comentários¹¹. Relatos como “Perdi a vontade de viver”, “O choro é o que me acompanha, a tristeza também, não sei o que fazer, lágrimas e lágrimas...”, “Depressão é orível, tenho vontade de chora toda hora...” e “Eu estou 24 horas na bad, só esperando a hora de morrer...” aparecem como principais comentários.

3.2. Aquisição dos dados

Para a formação do conjunto de dados, foi utilizada a *application programming interface* (API) disponibilizada pelo Youtube¹², com base em uma arquitetura *REST*, utilizando como estrutura de dados a representação em JSON¹³. Um *script* em Python foi desenvolvido para realizar a comunicação com a API e gerar os dados coletados em arquivo CSV.

⁸<https://www.youtube.com/>

⁹<https://www.billboard.com/articles/columns/pop/8488652/ariana-grande-thank-u-next-biggest-music-video-debut-youtube>

¹⁰<https://www.youtube.com/user/Itsallaboutmakeup14/featured>

¹¹<https://www.youtube.com/watch?v=K19Jhx9eX-Q>. Pesquisa realizada no dia 23 de março de 2019

¹²<https://developers.google.com/youtube/v3/>

¹³Javascript Object Notation

A *string* de busca utilizada na primeira requisição HTTPS foi “depressao+suicidio”, respeitando somente vídeos publicados para o Brasil. Também foi considerado um período limite de até três anos retroativos entre a data da publicação do vídeo e a data da pesquisa, o que resultou em um total de 26.908 vídeos relacionados, dispostos em registros paginados.

Para cada vídeo encontrado, iniciou-se o processo de iteração. Em cada etapa, uma nova requisição HTTPS foi realizada para a extração dos comentários do vídeo, cujo o resultado limitava-se a 100 entradas por página. Cada comentário extraído, era registrado imediatamente em um único arquivo CSV. Dados que pudessem identificar o autor, como *link* do comentário e nome, foram desprezados com o objetivo de respeitar sua privacidade. Ao final da página, uma nova requisição era submetida solicitando a próxima página, repetindo este processo até o fim dos registros.

Por restrições e técnicas desconhecidas, algumas requisições não retornavam os dados solicitados, sendo então desconsideradas da presente etapa. Testes prévios não indicaram um padrão frequente que indicasse aos autores o motivo das indisponibilidades. Dessa maneira, não foi possível efetuar uma busca para contornar o problema. No processo final, tais falhas foram computadas, resultando em uma perda de 34% do total dos comentários relacionados. Ao final de todo o processo, 312.055 comentários foram extraídos e formaram o conjunto de dados inicial.

Aladağ et al. [2018], em um trabalho publicado sobre detecção de idealizações suicidas em fóruns de Internet, consideraram em seus experimentos apenas os *posts* com 100 ou mais caracteres. De forma análoga, para o trabalho proposto neste artigo, foram selecionados somente comentários com 100 ou mais caracteres, resultando em um total de 13.182 comentários.

Em seguida, a tarefa de *case folding* (conversão dos tokens para letra minúscula) foi efetuada, bem como a remoção das pontuações. Para este estudo foram selecionados 718 entradas, sendo 376 entradas classificadas como “S - indícios de depressão” e 342 entradas como “N - sem indícios de depressão”, gerando assim o *dataset* para o experimento. As Tabelas 1 e 2 detalham o resultado final.

Tabela 1. Etapas de extração dos dados

Etapas de extração dos dados	Qtd. registros
1 - Comentários extraídos dos vídeos	312.055
2 - Comentários contendo 100 ou mais caracteres	13.182
3 - Comentários utilizados para o experimento	718

Tabela 2. Classificação dos comentários

Classe	Qtd. registros
S - início de depressão	376
N - sem início de depressão	342

3.3. Classificação do conteúdo depressivo

A classificação de comentários depressivos ocorreu com uma avaliação conjunta de 3 juízes, dentre eles, um psicólogo. Cada julgador recebeu uma cópia dos dados em uma

tabela Excel, contendo 718 entradas e com duas colunas: Uma exibia o comentário (*post*), e a outra reservada para classificação (*indício de depressão, S ou N*). Cada comentário foi lido e avaliado de forma independente, seguindo as características relatadas no manual DSM. Em seguida, os comentários que obtiveram consenso entre um avaliador e o psicólogo foram utilizados. A Tabela 3 apresenta alguns exemplos de comentários classificados pelos julgadores.

Tabela 3. Exemplos de comentários classificados pelos julgadores

Comentário	Indício de depressão (S/N)
<i>“Tenho 21 anos, sou diagnosticada com depressão, tomo remédios pro mesmo, faço tratamento com psiquiatra e psicólogo e sério nada disso adianta, minha vida é resumida em ficar deitada o dia inteiro, passo meu tempo dormindo pra não ter que ficar acordada, não tenho ânimo pra nada, nada me motiva, pensamentos a mil, sério não tenho mais prazer em viver minha vida tá se acabando aos poucos, e em breve vou estar em um lugar bem distante onde tudo isso que passo vai se acabar.”</i>	S
<i>“Eu sei que não estamos mais em setembro, mas gostaria de lembrar que ”setembro amarelo” é o mês da conscientização sobre a prevenção do suicídio e que esse movimento não tem tanta atenção quanto deveria. A depressão é, na maioria das vezes, o motivo desse tipo de acontecimento. Depressão é coisa séria, gente, e tem que ser tratada. Tem muita gente que precisa de ajuda, encorajamento e apoio por qualquer motivos que seja, e nós temos que olhar por eles. Pode ser um amigo nosso, um parente, namorado, ou sei lá. Alguém próximo de você pode estar passando por isso, então tente conversar, ajudar, e se manter saudável também. Não é demérito nenhum procurar ajuda. Até as mentes mais saudáveis podem precisar de terapia as vezes.”</i>	N
<i>“Minha família me odeia eu me corto tento me mata todos dizem que é frescura minha mas não é eu tenho depressão mas ninguém liga pra se eu estou viva ou morta a morte é a única esperança pra poder se liberta do sofrimento a vida é uma tortura viver é como se tivesse me matando de dentro pra fora isso eu não desejo pra ninguém nem pra que me faz tanto mal.”</i>	S
<i>“Nilce, muito obrigado por esse vídeo. Eu me lembro de quando eu comecei á assistir o cac. Tava lá pelo ep.400, e quem diria que, quando eu fosse diagnosticado com depressão, iria aparecer um vídeo sobre essa monstruosidade. Esse foi um vídeo que me emocionou muito, pois ele me lembra desses dias mais alegres. Tudo que eu posso dizer é, obrigado mesmo”</i>	N

A Tabela 4 detalha os sintomas que serviram de apoio aos jurados para avaliação e classificação dos comentários. É importante destacar que essa tabela foi enviada para os jurados. Ao final das avaliações, conforme mencionado, restaram 376 entradas classifica-

das como “S” (*há indícios de depressão*) e 342 entradas classificadas como “N” (*não há indícios de depressão*).

Tabela 4. Diagnóstico de depressão

Numero	Sintoma
1	Sentimentos depressivos a maior parte do dia ou quase todos os dias, conforme indicado por relatório (por exemplo, se sente triste, vazio, sem esperança) ou observação feita por outros
2	Perda notória de interesse ou prazer em todas ou quase todas as atividades diárias, em todo o dia ou quase todos os dias
3	Perda ou ganho de peso significativo sem dieta (por exemplo, uma mudança de mais de 5% do peso corporal em um mês), ou diminuir ou aumentar o apetite quase todos os dias
4	Insônia ou hipersonia quase todos os dias
5	Agitação ou retardo psicomotor quase todos os dias (observável por outros, não meros sentimentos subjetivos de inquietação ou de desaceleração)
6	Fadiga ou perda de energia quase todos os dias
7	Sentimentos de inutilidade ou culpa excessiva ou inadequada (que pode ser delirante) quase todos os dias
8	Diminuição da capacidade de pensar ou se concentrar, ou indecisão, quase todos os dias (seja por conta subjetiva ou como observado por outros)
9	Pensamentos recorrentes de morte (não apenas medo de morrer), ideação suicida recorrente sem plano específico, ou uma tentativa de suicídio ou um plano específico para cometer suicídio

4. Metodologia

A metodologia desse estudo utiliza como base um léxico contendo termos relacionados à depressão e a versão em português do LIWC (Linguistic Inquiry and Word Count) [Balage Filho et al., 2013], denominado nesse estudo LIWC2007-pt-br. A Seção 4.1 descreve brevemente o léxico depressivo utilizado. Em seguida, a Seção 4.2 descreve brevemente o LIWC2007-pt-br. Por fim, a Seção 4.3 descreve como os dois léxicos foram integrados para a criação de um novo léxico.

4.1. Léxico depressivo

O conjunto de dados utilizado neste trabalho é denominado *Depress-pr-br* [Nascimento et al., 2018], sendo este um léxico composto por 335 termos de cunho depressivo, que foi elaborado por meio da extração de conteúdos de redes sociais, em que tais conteúdos estão relacionados a manifestações de seus usuários contendo fortes indícios de depressão.

Com auxílio de outro léxico afetivo (ANEW-br [Kristensen et al., 2011]), o *Depress-pr-br* concentra de forma exclusiva um conjunto de palavras de valência negativa, pois segue a premissa de que indivíduos depressivos tendem a utilizar mais deste conjunto de palavras em seu vocabulário [Rude et al., 2004]. A Figura 1 mostra a nuvem de palavras selecionadas para a elaboração do léxico.

Tabela 5. Média F1 para os algoritmos de classificação NB, MNB, LMT, J48 e SVM na tarefa de detecção de sintomas depressivos em textos. Os valores em negrito indicam os melhores resultados com a utilização de cada algoritmo.

	NB	MNB	LMT	J48	SVM
LIWC_2007pt + Depress_br	0,775	0,789	0,795	0,726	0,804
LIWC_2007pt	0,775	0,793	0,795	0,722	0,796

Os algoritmos SVM (0,804) e J48 (0,726), apresentaram um melhor desempenho com a utilização do léxico. O algoritmo MNB (0,793), apresentou um melhor desempenho sem a utilização do léxico como apoio. Os algoritmos NB (0,775) e LMT (0,795), não apresentaram melhorias com a utilização do léxico.

A utilização de conteúdos com 100 ou mais caracteres revelou neste estudo conteúdos com indícios depressivos utilizando mais palavras tendem a ter uma maior correlação com os diagnósticos estabelecidos pelo [DSM-V, 2013] (vide a Tabela 4). Isso corrobora com um melhor diagnóstico depressivo, assemelhando-se a uma espécie de desafogo da doença ao expor todo o seu sentimento e pensamento, segundo a avaliação do psicólogo.

Conforme mencionado, alguns conteúdos foram descartados neste experimento por conterem menos de 100 caracteres (e.g., “*eu quero me matar*”, “*Só enxergo saída no suicídio*”). Apesar de demonstrarem indícios depressivos, estes conteúdos apresentam menos características quando comparados com conteúdos mais densos. Portanto, conclui-se que em tarefas de classificação de conteúdos depressivos, utilizar textos com uma grande quantidade de palavras contribui melhor para a identificação de sintomas da doença. A Tabela 6 apresenta alguns exemplos de conteúdos com 100 ou mais caracteres.

Muitos estudos sobre detecção de conteúdos com indícios depressivos em redes sociais, utilizam plataformas como Twitter e Facebook para a coleta de dados [Aldarwish and Ahmad, 2017; Chen et al., 2018; Tsubawa et al., 2015]. Neste estudo, a rede social Youtube se mostra como uma potencial plataforma para pesquisas do gênero.

O estímulo do conteúdo audiovisual do Youtube contribui para os usuários expressarem seus mais profundos sentimentos nos comentários. A interação audiovisual é mais adotada pelos usuários do que iterações restritas apenas a conteúdos em áudios e textos, pois desperta um sentimento mútuo entre espectador e apresentador [Mate and Curcio, 2009]. Como exemplo, um vídeo com o título “8 Sinais Que Indicam Que Você Está com Depressão e Não Sabe” dispõe de diversos comentários como “*chorei assistindo video, pois a muito tempo estou nessa situacao e ngm me compreende...*”, “*me identifico muito! nunca fui no medico por falta de grana mesmo! vivo numa angustia...*”, “*esse video mexeu muito comigo, sofro de depressao muitos anos...*”. Para trabalhos futuros, sugere-se o aprimoramento do léxico por meio da inclusão de novas palavras, bem como explorar os dados da plataforma Youtube, o que pode contribuir para uma melhor compreensão do comportamento social em pesquisas sobre depressão e muitas outras.

6. Conclusão

Os resultados desse trabalho indicam que a utilização de um léxico com termos de cunho depressivo pode melhorar a tarefa de detecção de sintomas depressivos em análise de

Tabela 6. Exemplos de conteúdos com 100 ou mais caracteres

<p><i>“Obrigado, hj eu tava na escola chorando mto, e mto triste. Pq acho que ninguém me ama, e ao ouvir essa música encontrei o caminho pra continuar lutando e superar tudo que passei esse ano, como término de namoro, ve a pessoa que amo com outro, ver meus pais brigando por n ter dinheiro pra pagar certa contas, e algumas vezes até fome... mas graças a essa música e a vc. Encontrei forças pra continuar a lutar mto obrigado sei que esse comentário vc nunca verá mas... eu queria te falar que vc fez a diferença e se possível me chama no whatsapp pra mim poder te agradecer melhor. Vlw msm”</i></p>
<p><i>“eu tenho depressão tudo sempre deu errado pra mim perde tudo que eu mais amava meus filhos meus irmãos minha mãe sou um não sou feliz em nada não conseguir termina meus estudos não consegui nem se quer ter uma profissão e ágora perde a unica pessoa que me ajudava minha mãe acabou pra mim tentei sai dessa depressão des dos meus 15 anos de idade ja tou com 32 anos pra não da mais pra luta e agora que eu perde minha mãe não aguento mais isso w terrível”</i></p>
<p><i>“Bom gente to aqui para desabafar Eu sofro a alguns anos de depressao eu só tenho 10 anos e minha mente já está suicida infelizmente. Eu nunca tive coragem de falar para minha família sobre isso porque minha irmã já sofre de depressao então acho que minha mãe não conseguiria aguentar pagar 2 psicólogos então só escondo tudo e eu sei que existem psicólogos de graça mas não sei se minha mãe suportaria duas filhas suicidas. Entãoo brigada a todos que leram esse comentário”</i></p>

textos. O uso de algoritmos utilizados em Aprendizagem de Máquina (e.g. *SVM*, *Naive Bayes*, *Logistic Model Tree*) quando combinado com outros recursos (neste estudo, o léxico), pode contribuir em tarefas de classificação.

Concluiu-se também que a rede social Youtube apresenta grande potencial para trabalhos de comportamento humano em redes sociais. Utilizar um processo automatizado para detectar um grande número de comentários de redes sociais com indícios de depressão pode contribuir em pesquisas psicológicas e de compreensão de comportamento social, auxiliando em grupos de apoio à depressão, seja no diagnóstico ou prevenção em grande escala.

A depressão é um problema de extrema relevância e que vem tomando proporções alarmantes. Esse problema precisa de atenção em programas de políticas públicas para diagnóstico e prevenção da doença, dessa maneira, este trabalho também contribuiu com a criação de um conjunto de dados, que pode servir como base para trabalhos que se dediquem a identificar sintomas depressivos em textos em português do Brasil.

Referências

- Aladağ, A. E., Muderrisoglu, S., Akbas, N. B., Zahmacioglu, O., and Bingol, H. O. (2018). Detecting suicidal ideation on forums: Proof-of-concept study. *Journal of Medical Internet Research*, 20(6).
- Aldarwish, M. M. and Ahmad, H. F. (2017). Predicting depression levels using social media posts. *2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS)*.
- Balage Filho, P. P., Pardo, T. A., and Aluísio, S. M. (2013). An evaluation of the Brazilian

- Portuguese LIWC dictionary for sentiment analysis. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology (STIL)*, pages 215–219.
- Bradley, M., L. P. and Cuthbert, B. (1999). Affective norms for english words (ANEW). *Gainsville, Fla, NIMH Centre for the Study of Emotion and Attention, University of Florida*.
- Bradley, M. M. and Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings.
- Chen, X., Sykora, M. D., Jackson, T. W., and Elayan, S. (2018). What about mood swings: Identifying depression on Twitter with temporal measures of emotions. In *Companion Proceedings of the The Web Conference 2018, WWW '18*, pages 1653–1660, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.
- DSM-IV (2011). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*, volume IV. American Psychiatric Association.
- DSM-V (2013). *Diagnostic and statistical manual of mental disorders*, volume V. American Psychiatric Publishing.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.
- IBGE (2013). Pesquisa nacional de saúde: 2013 : percepção do estado de saúde, estilos de vida e doenças crônicas: Brasil, grandes regiões e unidades da federação.
- IBGE (2017). Pesquisa nacional por amostra de domicílios contínua - pnad contínua.
- Kristensen, C. H., Gomes, C. F. A. d. A., Justo, A. R., and Vieira, K. (2011). Normas brasileiras para o Affective Norms for English Words. *Trends in Psychiatry and Psychotherapy*, 33:135 – 146.
- Mate, S. and Curcio, I. (2009). Mobile and interactive social television. *IEEE Communications Magazine*, 47(12):116–122.
- Nascimento, R., Cruz, P., dos Santos, G., and Guedes, G. P. (2018). Identificando sinais de comportamento depressivo em redes sociais. In *7th Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2018)*, Natal, RN, Brasil. SBC.
- Nguyen, T., Phung, D., Dao, B., Venkatesh, S., and Berk, M. (2014). Affective and content analysis of online depression communities. *IEEE Transactions on Affective Computing*, 5(3):217–226.
- Pennebaker, J. W. (2011). *The secret life of pronouns: what our words say about us*. Bloomsbury Press.
- PNAD, I. (2017). Acesso à internet e à televisão e posse de telefone móvel celular para uso pessoal : 2017.
- Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied psychological measurement*, 1(3):385–401.
- Rude, S., Gortner, E.-M., and Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8):1121–1133.
- Silveira, B., da Silva, A. P. C., and Murai, F. (2018). Análise de comunidades de suporte a transtornos de saúde mental do Reddit. In *7th Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2018)*, Natal, RN, Brasil. SBC.
- Tausczik, Y. R. and Pennebaker, J. W. (2009). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.

- Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., and Ohsaki, H. (2015). Recognizing depression from Twitter activity. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI 15*.
- WHO (2015). Comprehensive mental health action plan 2013–2020.
- WHO (2017). Depression.