

# DP-Symptom-Identifier: uma estratégia para classificar sintomas de depressão utilizando um conjunto de dados textuais na língua portuguesa

Vinicius Casani<sup>1</sup>, Alinne C. Correa Souza<sup>2</sup>, Rafael G. Mantovani<sup>3</sup>,  
Francisco Carlos M. Souza<sup>2</sup>

<sup>1</sup>ANYMARKET – Maringá – PR – Brasil

<sup>2</sup>Universidade Tecnológica Federal do Paraná – Dois Vizinhos – PR – Brasil

<sup>3</sup>Universidade Tecnológica Federal do Paraná – Apucarana – PR – Brasil

viniciuscasani3@gmail.com

{alinnesouza, rafaelmantovani, franciscosouza}@utfpr.edu.br

**Abstract.** *Depression is a psychological disorder that affects millions of people in the world, regardless of their age, social class, or nationality. Different techniques have been studied to analyze and recognize depression in the literature, such as Natural Language Processing, Sentiment Analysis. However, in Brazilian Portuguese, only a few studies have been focused on creating a dataset to classify symptoms of depression. In this paper, we propose a strategy called DP-Symptom-Identifier for collecting twitter messages and generate a novel text dataset with sentences that refer to the main symptoms of depression. Initial experiments with this dataset and different algorithms obtained high accurate performance values, which shows the research in this area is promising.*

**Resumo.** *A depressão é um distúrbio psicológico que afeta milhões de pessoas no mundo, indiferente à idade, classe social ou nacionalidade. Diferentes técnicas tem sido exploradas para analisar e reconhecer sintomas depressivos na literatura, como técnicas de Processamento de Linguagem Natural e Análise de Sentimentos. Entretanto, para o português brasileiro, poucos estudos tem proposto datasets para a classificação de sintomas da depressão. Neste artigo, propomos uma estratégia chamada DP-Symptom-Identifier para coletar tweets e criar um novo dataset com sentenças que possuem sintomas da depressão. Experimentos iniciais usando diferentes algoritmos obtiveram um alto desempenho preditivo, o que mostra que as pesquisas nesta área são promissoras.*

## 1. Introdução

Considerada o mal do século pela Organização Mundial da Saúde (OMS) [Organization et al. 2017], a depressão é um transtorno psiquiátrico que afeta o emocional da pessoa. Segundo a OMS, a depressão atinge pessoas de todas as idades, classes sociais e nacionalidades. Além disso, o número de diagnósticos aumentou em 18,4% entre 2005 e 2015, afetando 4,4% da população mundial. No Brasil, a parcela da população afetada é de 5,8%, o que coloca o país como o maior detentor do transtorno na América Latina, e o segundo nas Américas, ficando atrás apenas dos Estados Unidos, que tem 5,9% da sua população afetada.

Nesse contexto, técnicas de Análise de Sentimentos (AS) têm sido amplamente utilizadas para detecção e extração de sentimentos a partir de dados textuais. Historicamente a análise de textos é uma tarefa complexa, custosa e tediosa quando realizada manualmente. No entanto, com o atual avanço da tecnologia e o desenvolvimento de áreas como Aprendizado de Máquina (AM) e Processamento de Linguagem Natural (PLN); e uma grande quantidade e disponibilidade de dados gerados por redes sociais, fóruns e páginas na Web, é possível extrair informações riquíssimas e relevantes de dados textuais.

Dentre as vertentes de trabalhos em AS, alguns em específico tem explorado a detecção automática e textual de distúrbios mentais [Hassan et al. 2017, Islam et al. 2018]. Entretanto, esta é ainda uma área de pesquisa emergente, principalmente no Brasil, onde apenas um trabalho relacionado pode ser encontrado [Rosa et al. 2019]. Tendo em vista essa lacuna na literatura, este trabalho tem como objetivo apresentar uma estratégia denominada de *DP-Symptom-Identifier*, a qual foi criada para coletar e identificar os sintomas depressivos em postagens do *Twitter* utilizando técnicas de Processamento de Linguagem Natural (PLN). Para isso, foi desenvolvida uma aplicação Web para coleta, análise e rotulação, indicando as respectivas categorias de sintomas depressivos, como: comportamental, fisiológico e/ou psíquico. Esse processo de rotulação foi acompanhado por uma psicóloga especialista no domínio.

O restante do artigo está estruturado da seguinte forma. Na Seção 2 são descritos os trabalhos relacionados; na Seção 3 é detalhada a estratégia *DP-Symptom-Identifier*; na Seção 4 são apresentados alguns resultados iniciais usando o conjunto de dados gerado, e por fim, na Seção 5 são apresentadas as conclusões e trabalhos futuros.

## 2. Trabalhos Relacionados

Nos últimos anos, as técnicas de AS [Liu 2012] têm recebido atenção especial devido ao seu forte potencial para extrair sentimentos de dados textuais e, conseqüentemente, auxiliar no diagnóstico de perfis depressivos. Neste contexto, alguns estudos [Birjali et al. 2017, Ma et al. 2017, Aldarwish and Ahmad 2017, Keumhee Kang et al. 2016] utilizaram o *Twitter* como fonte de dados para a criação de base de dados com postagens em inglês. Os dois primeiros estudos realizaram a rotulação dos dados de forma automática, enquanto os demais de forma manual.

Em [Birjali et al. 2017] os autores optaram por buscar postagens do *Twitter* e as rotularam de forma automática utilizando o método de similaridade semântica *WordNet*, tendo como base termos associados com a depressão. Já em [Ma et al. 2017] os autores criaram a base de dados baseado em postagens do *Twitter* contendo a palavra “depressão”; postagens de contas de profissionais de saúde mental, e de *blogs* sobre depressão. No final do processo, a base de dados compilou um total de 54 milhões de postagens.

No que diz respeito a rotulação manual, no estudo de [Aldarwish and Ahmad 2017] os autores criaram um dataset com 2354 postagens do *twitter*, 2132 do *LiveJournal* e 2287 do *Facebook*. Deste total, 2073 postagens foram rotuladas como deprimidas e 4700 como não deprimidas. Além disso, as postagens identificadas como deprimidas foram sub-categorizadas entre os nove sintomas definidos pelo DSM-5<sup>1</sup>. Por fim, no trabalho de [Keumhee Kang et al. 2016] foram coletadas

---

<sup>1</sup>Diagnostic and statistical manual of mental disorders

postagens do *Twitter* que continham as seguintes palavras ou sentenças: “*Christmas*”, “*Suicide*”, “*I feel relaxed*”, “*I feel good*”, “*want to die*”, “*I feel stressed*”, “*I feel sad*”, “*kill myself*” e “*want to commit suicide*”). Para cada postagem era possível aplicar os seguintes rótulos: negativo, neutro ou positivo.

É importante destacar que apesar do número considerável de estudos, somente o estudo de [Rosa et al. 2019] criou uma base de dados para a língua portuguesa, como também destacado no mapeamento sistemático anteriormente realizado [Casani et al. 2021]. Neste estudo em específico foram filtradas sentenças de postagens do *Facebook* que continham expressões como “odeio minha vida”, “me sentindo triste”, “estou estressado”, entre outras. O estudo também utilizou um conjunto de sentenças positivas para classificar como “não depressão”. Entretanto, considerando todos esses estudos, é possível identificar um mesmo problema na criação das bases de dados. A maioria dos estudos concentra-se na identificação de sentimentos positivos e negativos, porém a resolução seguindo uma classificação binária não é suficiente para identificar os sintomas depressivos. Desta forma, em comparação aos trabalhos descritos, o principal diferencial neste artigo é a exploração de uma análise considerando três categorias de textos depressivos na língua portuguesa: sintomas psíquicos, comportamentais e fisiológicos. Segundo o DSM-5 [APA 2013], os sintomas psíquicos e comportamentais são mais relevantes para indicar um quadro depressivo do que os fisiológicos.

### 3. Estratégia *DP-Symptom-Identifier*

A estratégia *DP-Symptom-Identifier* foi criada com o objetivo de coletar mensagens compartilhadas no *Twitter* utilizando técnicas de PLN. A visão geral da estratégia é apresentada na Figura 1, a qual é composta por cinco etapas: (1) Coleta e rotulação dos dados; (2) Pré-processamento; (3) Extração das características.

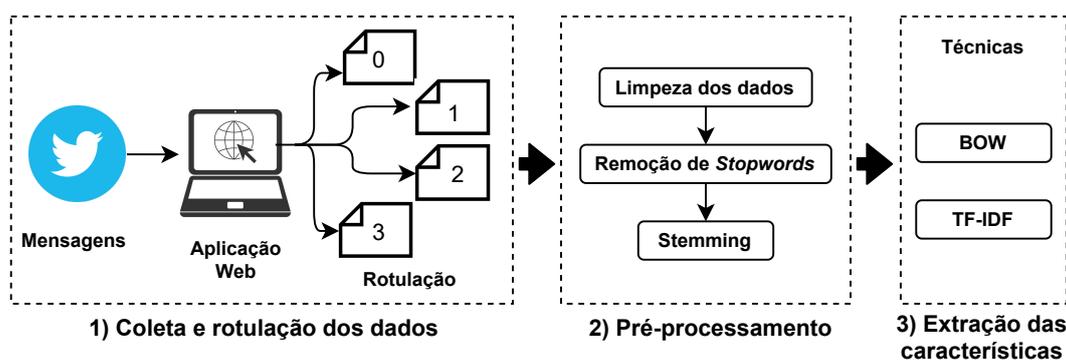


Figura 1. Estratégia *DP-Symptom-Identifier*.

A etapa 1 foi dividida em duas sub-etapas: (1.1) coleta dos dados; e (1.2) rotulação dos dados coletados. Na primeira sub-etapa (1.1) foi utilizada a API do *Twitter* para coletar as mensagens; e a segunda (1.2) consiste na rotulação dos dados, por meio de uma aplicação web, classificada em quatro classes possíveis: 0 - Nenhum, 1 - Fisiológico, 2 - Comportamental e 3 - Psíquico. A etapa 2 consiste no pré-processamento dos dados coletados, a fim de reduzir o tamanho da base de dados, bem como remover dados redundantes e irrelevantes utilizando as técnicas de *stopwords* e *stemming*. A etapa 3 visa extrair as características dos dados textuais da base utilizando as técnicas *Bag-of-Words* (*BoW*) e *Term Frequency-Inverse Document Frequency* (*TF-IDF*).

### 3.1. Coleta e rotulação dos dados

Para realizar a coleta inicialmente foi criada uma base de dados com o auxílio de uma psicóloga para treinar o modelo. Essa base contém 200 sentenças<sup>2</sup> que remetem a sinais sintomáticos da depressão, rotuladas em três categorias: (i) comportamental, (ii) psíquico e (iii) fisiológico. Alguns exemplos de sentenças e seus respectivos rótulos são apresentadas na Tabela 1. É importante salientar que o rótulo dos exemplos pode apresentar quaisquer uma das categorias sintomáticas: Fisiológico, Comportamental, Psíquico, ou ausência completa de sintomas (Nenhum).

**Tabela 1. Sintomas presentes em sentenças rotuladas com ajuda da psicóloga.**

Sentença	Categoria Sintomática
Eu quero morrer	Comportamental
Não desejo sair de casa	Comportamental
Tenho dificuldade para dormir	Fisiológico
Estou sempre cansado	Fisiológico
Me sinto inútil	Psíquico
Sou infeliz	Psíquico

Para auxiliar a coleta das mensagens foi desenvolvida uma ferramenta em Java, denominada Aplicação de Coleta de Dados (ACD), que utiliza a API do Twitter para buscar mensagens compartilhadas na rede social de acordo com palavras-chave. As funcionalidades implementadas na ACD possibilitaram utilizar palavras e sentenças como parâmetros de filtragem durante a seleção dos *tweets*. Neste contexto, somente postagens que continham os parâmetros desejados eram retornados, como por exemplos as sentenças apresentadas na Tabela 1. Dentre os *tweets* retornados, foram excluídos os que representavam um *retweet*<sup>3</sup>.

Uma vez os *tweets* coletados e salvos em disco, os mesmos foram rotulados. Para isso, foi desenvolvida uma segunda Aplicação Web (AW) para que a psicóloga e especialista no domínio definisse os rótulos das postagens de acordo com as categorias sintomáticas apresentadas na Tabela 1. Na Figura 2 é ilustrada a tela referente a análise das rotulações realizadas. Nesta tela, as postagens apresentadas são limitadas a um mesmo usuário, o qual pode ser filtrado pelo campo de texto disponível. Os registros apresentados são ordenados por data da publicação de forma decrescente, simulando uma *timeline* da atividade do usuário ao longo do tempo para facilitar a análise.

A Figura 3 exibe a tela para facilitar o processo de rotulação dos dados coletados. Por meio desta tela é possível rotular cinco postagens por vez, marcando ou desmarcando a opção para cada um dos sintomas. Após a rotulação de todas as postagens, o botão presente no canto inferior direito é acionado para gravar as informações alteradas.

Após o processo de rotulação, cada postagem pode apresentar os seguintes valores para rótulo: (i) **Nenhum**: quando a postagem não possui nenhum dos sintomas;

<sup>2</sup>Essa base pode ser acessada no link: <https://github.com/fcarlosmonteiro/dp-symptom-identifier>

<sup>3</sup>*Retweet* é uma republicação de um *Tweet*.

Dataset    Categorizar    Timelines

Usuário

Buscar Timeline

Texto	Classificação	Validado Especialista	Data
Porque eu tenho tanta insegurança nessa 'relação'?	Nenhum	✓	2019-10-28
Vou fazer um brigadeiro pra comer sozinha porque eu sou a minha melhor companhia	Psíquico	✓	2019-10-09
Eu tava disposta a fazer diferente dessa vez, caralho eu falei de nós até pra minha mãe..	Nenhum	✓	2019-08-11
Sobre os relatórios <a href="https://t.co/kZ7MShCNju">https://t.co/kZ7MShCNju</a>	Psíquico	✓	2019-11-01
O comentário que recebi hoje: "Sorte de que tá 'comendo'. Eu: verdade, um total de 0 pessoa Mereço mesmo kkkkkkkkkkkkkk	Nenhum	✓	2019-10-26

Figura 2. Tela da aplicação AW referente a análise dos dados.

(ii) **Psíquico**: quando a postagem possui apenas indícios de sintomas psíquicos; (iii) **Comportamental**: nos casos em que a postagem possui indicação de evidências comportamentais; e (iv) **Fisiológico**: quando o sintoma identificado na postagem é de origem fisiológica. Esses valores foram definidos juntamente com a psicóloga e com base nos principais sintomas da depressão. Para a avaliação apropriada foram criadas duas bases de dados distintas: (i) Base de treinamento; e (ii) Base de dados de testes. A base de dados de treinamento foi criada utilizando as postagens de 1988 usuários coletadas e rotuladas conforme o processo descrito, as quais deveriam estar em conformidades com o padrão das sentenças apresentadas na Tabela 1. Esta base possui um total de 2008 linhas e 7 coluna<sup>4</sup>. A Figura 4 demonstra um exemplo desta base.

Além do dado textual extraído da postagem “tweet”, as demais colunas descritivas são: “idTweet” coluna utilizada para identificação dos *tweets*, “fisiologico” coluna a qual remete se o *tweet* pertence a categoria de sintomas fisiológicos. A coluna “comportamental” é utilizada para controlar o fato do *tweet* ser ou não um sintoma da categoria comportamental. Já a coluna “psiquico” demonstra a presença ou não de sintomas psíquicos. É importante ressaltar que o mesmo *tweet* pode ser classificado em mais de uma categoria. A coluna “label” contém informações referente a rotulação do dado. A coluna “dhTweet” apresenta a data da postagem. Por fim a coluna “validadoEspec” informa se a postagem foi validada pela especialista que acompanha este trabalho.

Já a base de dados de testes foi criada utilizando a funcionalidade de busca de *timelines* por ID de usuário, existente na ACD. Esta base possui as mesmas colunas que

<sup>4</sup>Essa base pode ser acessada no link: <https://github.com/fcarlosmonteiro/dp-symptom-identifier>

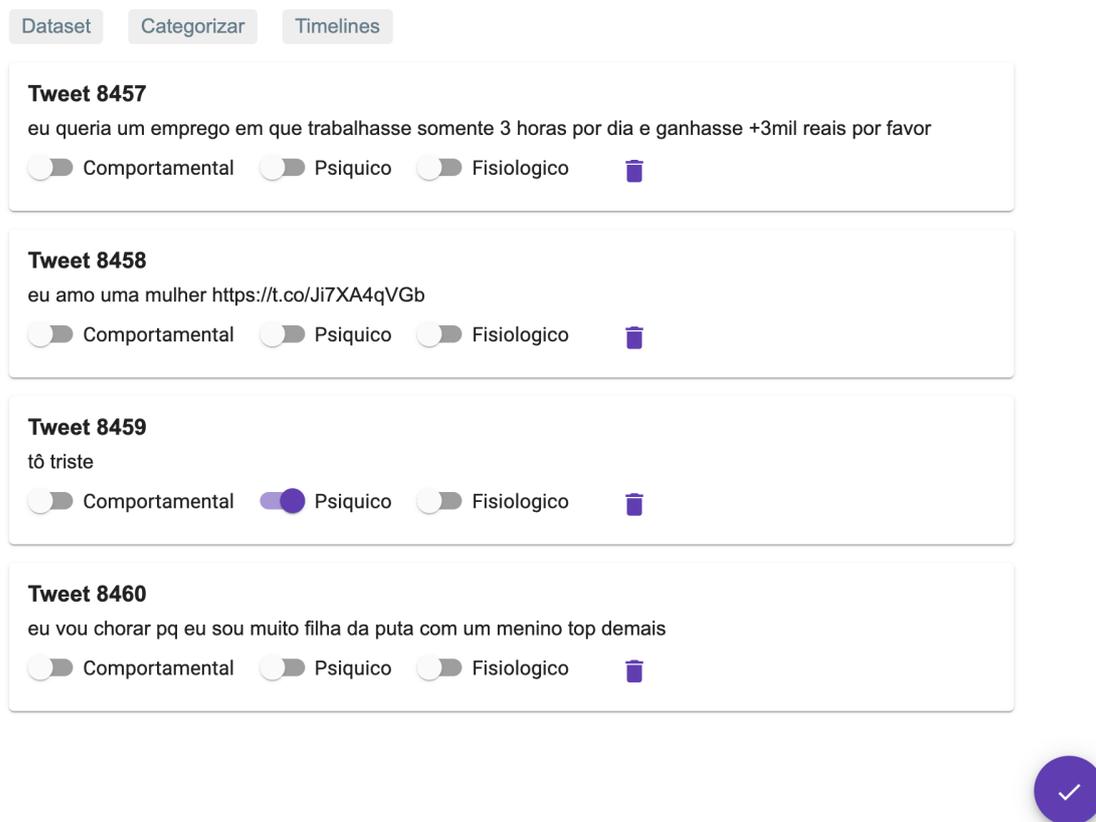


Figura 3. Tela da aplicação AW referente a rotulação dos dados.

tweet	idTweet	fisiologico	comportamental	psiquico	label	dhTweet	validadoEspec
1 só eu que já fiquei com dúvida se só tô muito triste o...	11	FALSE	FALSE	TRUE	3	2019-09-12	TRUE
2 Crises de ansiedade Depressão Tô me afundando e n ...	20	FALSE	FALSE	FALSE	0	2019-09-11	TRUE
3 hj vai passar uma materia sobre depressao no globo r...	22	FALSE	FALSE	FALSE	0	2019-09-11	TRUE
4 Hoje o Profissão Repórter fala sobre Depressão. ASSI...	23	FALSE	FALSE	FALSE	0	2019-09-11	TRUE

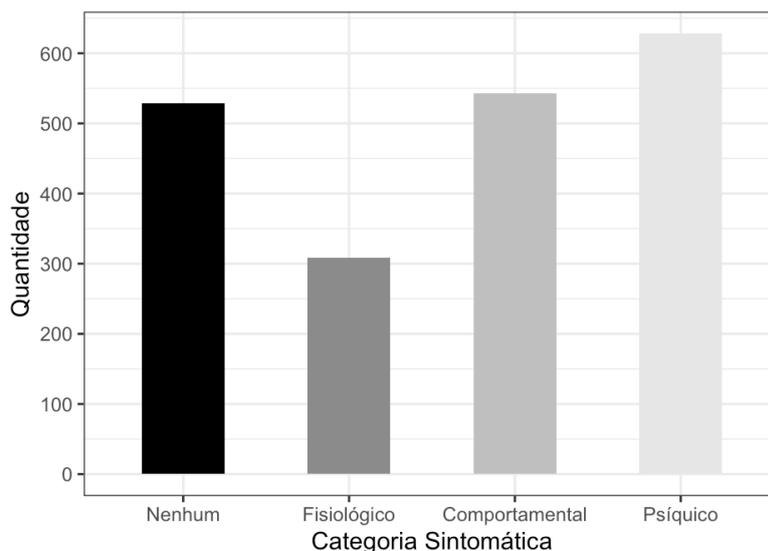
Figura 4. Exemplo da base de dados de Treinamento.

a base de dados de treinamento. Os dados armazenados nesta base podem ser utilizados para testar modelos preditivos quando fazendo o uso de algoritmos de AM, para classificar os dados. Neste contexto, a quantidade de subconjuntos ( $k$ ) escolhida para validar os modelos criados é de  $k = 10$  subconjuntos, valor frequentemente utilizado pela literatura.

Por fim, a Figura 5 demonstra a distribuição das classes existente na base de dados de treinamento. A classe que representa os sintomas psíquicos é a que possui a maior quantidade de exemplos dentre as demais. Isso pode ser justificado pelo fato de que esta categoria sintomática é a que possui uma maior quantidade de exemplos no *American Psychiatric Association* [APA 2013], e ainda, alguns sintomas da categoria comportamental são relacionados com os da categoria psíquico.

### 3.2. Pré-processamento

Esta etapa é realizada com a finalidade de reduzir o tamanho da base de dados, e também remover dados redundantes e irrelevantes. Em uma primeira etapa todos os caracteres das postagens foram convertidos para caracteres minúsculos. Em seguida, caracteres



**Figura 5. Distribuição das classes presentes no conjunto de treinamento.**

numéricos foram removidos, assim como pontuações, *links* e espaços desnecessários. Além disso, citações de outros usuários e *retweets* também foram removidos. Um último passo removeu todos os caracteres especiais, tais como *emojis*, pois estes não representam evidências sintomáticas da depressão. Em seguida, duas técnicas de PLN foram utilizadas: remoção de *stopwords* - palavras que devem ser desconsideradas na análise pois não apresentam informações relevantes para a construção do modelo que será analisado, como por exemplo "o", "e", "a", "de", "que"; e *stemming* para a redução das palavras ou termos ao seu radical.

### 3.3. Extração de Características

Para identificar a melhor forma de extrair as características dos dados textuais da base, foram realizados testes com duas técnicas: (i) *(TF-IDF)* e (ii) *BoW*. A técnica TF-IDF visa expressar a relevância de uma palavra em um dado *corpus*, a qual foi utilizada com o auxílio da função `TfIdfVectorizer` do pacote `superml`<sup>5</sup> do R. Ao executar a vetorização dos 2008 exemplos presentes na base de dados, o processo retornava uma matriz com mais de 10 mil atributos descritivos (colunas). Para tentar diminuir esta quantidade, foram identificados e removidos todos os atributos correlacionados, em uma porcentagem igual ou superior a 95%. Mesmo após realizar esta operação de redução, ainda ocorria a geração de um *dataset* longitudinal, no qual a quantidade de atributos (colunas) é maior que a quantidade de exemplos (linhas) disponível na base de dados. Nos primeiros testes realizados, tais bases longitudinais geravam modelos preditivos com *overfitting*: eles prediziam sempre a mesma classe (majoritária) e erravam a predição de todas as outras classes.

Por conta desses resultados uma segunda técnica foi utilizada, o *Bag-of-Words* (BoW), que consiste em a frequência que cada termo se repete dentro da base de dados. Usando o BoW a quantidade de atributos descritivos gerados foi de 4650. Após o processamento principal da técnica, foram removidos os atributos com uma frequência menor

<sup>5</sup><https://www.rdocumentation.org/packages/superml/versions/0.4.0>

que 10, reduzindo a quantidade total de atributos descritivos para 340. Tal mudança foi muito significativa, pois foi possível induzir modelos mais precisos, obtendo resultados superiores em até 50% quando comparado com os resultados obtidos com *TF-IDF*. Sendo assim, o *BoW* foi a técnica newadotada para realizar a extração de características dos dados textuais.

#### 4. Experimentos iniciais para avaliação do conjunto de dados

Com a base criada, um experimento pequeno foi realizado no intuito de avaliar a qualidade do conjunto de dados. O *setup* experimental definido utilizou apenas o conjunto de treinamento descrito nas seções anteriores, e quatro diferentes algoritmos de classificação: *Support Vector Machines* (SVMs), *Näive Bayes* (NB), *Multilayer Perceptron* (MLP) e *Random Forest* (RF). Os algoritmos foram executados por meio de uma validação cruzada usando 10 partições de dados estratificadamente amostrados. Foram criadas múltiplas tarefas preditivas binárias para cada categoria sintomática. O desempenho dos modelos induzidos foi mensurado por meio da AUC<sup>6</sup>, e realizada uma média para cada um dos algoritmos. Os resultados gerais são apresentados na Tabela 2.

**Tabela 2. Valores de AUC obtidos pelos algoritmos avaliados no conjunto de treinamento gerado pela estratégia**

Algoritmo	AUC	sd
RF	0.935	0.015
SVM	0.927	0.013
MLP	0.900	0.009
NB	0.725	0.020

Os resultados iniciais obtidos sugerem que a estratégia elaborada é descritiva e efetiva, pois as características textuais extraídas são mapeadas corretamente para as categorias sintomáticas da depressão. Isso é visivelmente corroborado pelos valores médios de AUC acima de 0.9, obtidos pelos algoritmos RF, SVM e MLP.

#### 5. Considerações finais

Neste trabalho foi desenvolvida a estratégia *DP-Symptom-Identifier* para coletar e analisar mensagens compartilhadas em português no *Twitter* a fim de identificar um dos três tipos de sintomas da depressão (Psíquicos, Fisiológicos e Comportamentais). A estratégia é composta por uma aplicação web que facilita o processo de coleta das postagens do *Twitter* e a rotulação dos dados, a qual foi realizada com o auxílio de uma psicóloga. Além disso, foi possível analisar as postagens com um experimento inicial, e comprovar a efetividade da proposta. Como trabalhos futuros pretense-se: expandir o conjunto de dados, adicionando mais *tweets*; rotula-los usando as ferramentas desenvolvidas; e incrementar os experimentos com os algoritmos preditivos. Deseja-se também expandir a coleta de dados para outras redes sociais.

<sup>6</sup>É uma medida de desempenho utilizada em problemas de classificação, que representa a medida de separabilidade, ou seja, mostra até que ponto o modelo avaliado é capaz de separar corretamente as classes [Flach et al. 2011].

## Referências

- Aldarwish, M. M. and Ahmad, H. F. (2017). Predicting depression levels using social media posts. In *2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS)*, pages 277–280.
- APA (2013). *Diagnostic and statistical manual of mental disorders : DSM-5*. American Psychiatric Association Arlington, VA, 5th ed. edition.
- Birjali, M., Beni-Hssane, A., and Erritali, M. (2017). A method proposed for estimating depressed feeling tendencies of social media users utilizing their data. pages 413–420. Springer International Publishing.
- Casani, V., Mantovani, R. G., Souza, A. C. C., and Souza, F. C. M. (2021). Identificação de perfis depressivos em redes sociais utilizando aprendizado de máquina: um mapeamento sistemático. *Computer on The Beach*, (12).
- Flach, P., Hernandez-Orallo, J., and Ferri, C. (2011). A coherent interpretation of auc as a measure of aggregated classification performance. pages 657–664.
- Hassan, A. U., Hussain, J., Hussain, M., Sadiq, M., and Lee, S. (2017). Sentiment analysis of social networking sites (sns) data using machine learning approach for the measurement of depression. In *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 138–140.
- Islam, M. R., Kamal, A. R. M., Sultana, N., Islam, R., Moni, M. A., and ulhaq, A. (2018). Detecting depression using k-nearest neighbors (knn) classification technique. In *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, pages 1–4.
- Keumhee Kang, Chanhee Yoon, and Eun Yi Kim (2016). Identifying depressive users in twitter using multimodal analysis. In *2016 International Conference on Big Data and Smart Computing (BigComp)*, pages 231–238.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.
- Ma, L., Wang, Z., and Zhang, Y. (2017). Extracting depression symptoms from social networks and web blogs via text mining. pages 325–330. Springer International Publishing.
- Organization, W. H. et al. (2017). Depression and other common mental disorders: global health estimates. Technical report, World Health Organization.
- Rosa, R. L., Schwartz, G. M., Ruggiero, W. V., and Rodríguez, D. Z. (2019). A knowledge-based recommendation system that includes sentiment analysis and deep learning. *IEEE Transactions on Industrial Informatics*, 15(4):2124–2135.