

Aplicação de Processamento de Linguagem Natural na identificação de perfis depressivos em mídias sociais

Application of Natural Language Processing to identify depressive profiles on social media

Beatriz Gonçalves Fontes¹, Flávio Rosendo da Silva Oliveira¹

¹Análise e Desenvolvimento de Sistemas - Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco – Campus Paulista (IFPE) - Paulista - PE - Brazil

bgf2@discente.ifpe.edu.br, flavio.oliveira@paulista.ifpe.edu.br

Resumo. *É perceptível a influência das mídias sociais na saúde mental das pessoas. Com o avanço tecnológico e a prevalência das interações virtuais, as relações sociais passaram a ser mediadas por dispositivos eletrônicos, suscitando preocupações quanto aos impactos psicológicos dessas plataformas. Este trabalho propõe a aplicação de técnicas de Processamento de Linguagem Natural para identificar possíveis perfis depressivos. Foram exploradas 5 técnicas de classificação consolidadas na literatura para compreender as interações textuais online, são elas: Logistic Regression; Bernoulli Naive Bayes; Random Forest; Stochastic Gradient Descent e Multilayer Perceptron. E dois métodos de pré-processamento textual: word2vec e Term Frequency-Inverse Document Frequencies. Ao utilizar técnicas como logistic regression e multilayer perceptron, os resultados obtidos revelaram-se promissores. Ambos os algoritmos alcançaram um F1-score de 79% utilizando a técnica de word embedding Term Frequency-Inverse Document Frequencies, indicando uma eficiência na identificação de possíveis perfis depressivos. Esses resultados mostram como tais tecnologias podem ser utilizadas como uma possível alternativa de apoio a iniciativas de bem estar social, com enfoque neste tema.*

Palavras-chave: *Mídias Sociais; Saúde Mental; Depressão; Análise de Sentimento; Processamento de Linguagem Natural.*

Abstract. *The influence of social media on people's mental health is noticeable. With technological advancement and the prevalence of virtual interactions, social relationships began to be mediated by electronic devices, raising concerns about the psychological impacts of these platforms. This work proposes the application of Natural Language Processing techniques to identify possible depressive profiles. 5 classification techniques consolidated in the literature were explored to understand online textual interactions, they are: Logistic Regression; Bernoulli Naive Bayes; Random Forest; Stochastic Gradient Descent and Multilayer Perceptron. And two textual preprocessing methods: word2vec and Term Frequency-Inverse Document Frequencies. When using techniques such as logistic regression and multilayer perceptron, the results obtained were promising. Both algorithms achieved an F1-score of 79% using the Term Frequency-Inverse Document Frequencies word embedding technique, indicating efficiency in identifying possible depressive profiles. These results show how such technologies can be used as a possible alternative to support social welfare initiatives, with a focus on this topic.*

Keywords: *Social Media; Mental Health; Depression; Sentiment Analysis; Natural Language Processing.*

1. Introdução

Com o avanço tecnológico na comunicação e o aumento do uso das mídias sociais, as pessoas passaram a utilizar mais seus dispositivos eletrônicos (Silva *et al.*, 2021), resultando em relações predominantemente virtuais. Isso reflete a realidade contemporânea, onde a interação social é mediada pela tela eletrônica, levando a uma desmaterialização do mundo social, conforme discutido por Dupas (2001, p. 16).

A ansiedade e depressão são algumas das doenças que mais afetam a população (Piga; Shima; Romanichen, 2021). Considerada o “Mal do século” pela Organização Mundial de Saúde (OMS), a depressão é usada para descrever um sintoma relacionado a patologias distintas, além de um estado emocional normal que afeta um indivíduo (Quemel *et al.*, 2021). A literatura mostra que a associação entre suicídio e transtornos mentais é de mais de 90%. Entre os transtornos mentais associados ao suicídio, a depressão maior se destaca (McGirr, A. *et al.*, 2007).

De acordo com (Liu, 2012) a Análise de Sentimento (AS), também conhecida como Mineração de Opiniões, é o campo de estudo que tem por objetivo analisar opiniões, sentimentos, avaliações, atitudes e emoções das pessoas, referente a serviços, organizações, eventos, produtos ou a figuras públicas. A AS tem sido frequentemente utilizada principalmente por empresas que procuram verificar a opinião dos clientes quanto aos seus produtos (Carmo Neto, 2022).

Em (Mendes, 2021), a identificação de indivíduos com o um Perfil Possivelmente Depressivo (PPD) contribui para intervenções e acompanhamento por profissionais da área da saúde focados nesse tema, desta forma sendo essencial para o direcionamento efetivo de recursos de saúde mental.

Uma pessoa com um quadro de PPD, quando diagnosticado com antecedência, possui mais chances de reverter seu quadro com menos dificuldade, se comparada com um indivíduo que descobre seu estado em um estágio avançado. Quando analisados separadamente, os sintomas da depressão tornam-se muito difíceis de serem diagnosticados, visto que muitas vezes seus sintomas passam despercebidos pelas pessoas (Júnior *et al.*, 2015). Nesse contexto, as Mídias Sociais que permeiam a internet podem servir de grandes aliadas. Estudos como os de (Choudhury *et al.*, 2013) mostram que as mídias sociais online são um dos melhores ambientes para identificação de PPD, quando comparados com técnicas usualmente utilizadas na área de saúde.

Na literatura, existem diversos exemplos bem-sucedidos de técnicas de Processamento de Linguagem Natural (PLN) que são aplicadas para saúde mental, como: a previsão de perfis com características depressivas no *twitter* através de uma estratégia criada pelos autores (Casani *et al.*, 2021), a identificação da existência de uma possível diferença entre o comportamento de usuários potencialmente depressivos e os usuários saudáveis de forma online (Vedula e Parthasarathy, 2017) e a detecção de intencionalidades suicidas em mensagens e notas textuais (Boukil *et al.*, 2019).

Este trabalho tem por finalidade contribuir a partir da automação da análise desses dados oriundos de Mídias Sociais através de PLN, já que a análise manual se torna inviável devido à grande quantidade de informações a serem processadas. Com isso, busca-se proporcionar *insights* a respeito do estado mental/emocional de indivíduos, a fim de identificar de forma precoce e eficaz um PPD. Tudo isso,

permitindo que esses *insights* possam vir a ser utilizados em estudos e projetos futuros, que objetivam melhorar o bem estar social.

Os trabalhos nesta área de pesquisa têm sido, conforme evidenciado na seção de trabalhos relacionados, conduzidos de maneira fragmentada, geralmente focando em um único algoritmo inteligente ou apresentando uma abordagem para aplicar PLN no contexto do problema abordado (i.e. PPD e condições análogas). Este trabalho busca preencher essa lacuna, explorando o uso de diferentes tipos de algoritmos inteligentes e dois métodos de pré-processamento textual, a fim de fortalecer a compreensão de como o domínio de PLN e AS podem dar apoio nesta área de estudo, identificando postagens de usuários com PPD.

As seções subsequentes deste documento são divididas em quatro partes distintas: Referencial Teórico, Metodologia, Experimentos e Resultados, e Conclusão. No Referencial Teórico, introduz-se como a análise de sentimentos está relacionada ao processo de compreensão e interpretação de opiniões, além da exploração de trabalhos afins à temática deste artigo. A seção de Metodologia descreve o método proposto neste estudo, apresentando a base de dados selecionada, e a configuração adotada para o treinamento dos modelos. Na seção de Experimentos e Resultados, os detalhes acerca da experimentação e seus resultados são apresentados. Por fim, na seção de Conclusão, realiza-se um resumo abarcando as realizações deste artigo, seguido de uma análise dos resultados obtidos, além da exposição de possíveis direções para trabalhos futuros.

2. Trabalhos relacionados

A comunicação e a tecnologia andam juntas, pois uma não iria avançar sem a outra. O virtual pode ser considerado a evolução da comunicação (Lévy, 2017). A comunicação, com a evolução da sociedade, passou a ter uma importância cada vez maior devido a grande quantidade de informações circulando no meio digital.

Pensando nesse cenário, muitas organizações e empresas passaram a expandir o seu mercado para o meio digital visando um ganho maior de consumidores, e conseqüentemente um faturamento ainda maior. Com isso, o acesso a produtos e serviços ficou cada vez mais facilitado, e por tanto, a procura por *feedbacks* e opiniões públicas — tanto por parte das empresas e organizações, quanto por parte dos próprios clientes — a respeito desses mesmos produtos e serviços não passou despercebida. A AS está completamente ligada a esse recolhimento de opiniões públicas (Carmo Neto, 2022).

Durante a realização dos estudos acerca do tema proposto neste trabalho, foram encontrados alguns que se destacaram. O primeiro consiste em um mapeamento sistemático (Casani *et al.*, 2021) publicado em 2019 no Repositório Institucional da Universidade Tecnológica Federal do Paraná (RIUT). O segundo foi um mapeamento (Johannben; Biemann, 2018) publicado no SpringerLink no ano de 2019.

Casani (2021) criou uma estratégia intitulada de *PD-Symptom-Identifier*, com o propósito de auxiliar no processo de localizar/identificar perfis com características depressivas no *Twitter*, e que posteriormente possibilita a análise do perfil como um todo. O estudo contou com o auxílio de uma especialista em psicologia, que contribuiu na análise dos dados e resultados, para identificar quais usuários apresentavam pelo menos um dos seguintes três quadros de tipos de sintomas de depressão: Psíquicos, Fisiológicos e Evidências Comportamentais. Para chegar nos resultados desejados, o

estudo apresentou a utilização de técnicas de PLN, e algoritmos de classificação de padrões. Nesse caso o algoritmo selecionado foi o *Multilayer Perceptron* (MLP), que teve uma taxa de acurácia de 90% com um tempo de processamento inferior a 5 segundos, provando que a estratégia proposta para identificar perfis depressivos nas *timelines* do *Twitter* conseguiu cumprir seus objetivos.

Em (Johannben; Biemann, 2018), o autor mostra uma revisão de estudos que possuem relação com técnicas e algoritmos de mineração de dados em problemas de bem-estar mental. O foco desse estudo se apresentou bastante amplo, pois além da depressão o mesmo abrangeu outros problemas de saúde mental em geral. Entre os resultados que tinham como foco a depressão, os autores identificaram que os algoritmos de *Support Vector Machines* (SVM) e *Naive Bayes* (NB) foram as abordagens mais utilizadas em estudos com esse tipo de linha de pesquisa.

Diferentes estudos utilizaram o *Twitter* como fonte de dados para a construção de suas bases de dados. Dentre tais estudos, é coerente ressaltar a importância da pesquisa de (Vedula e Parthasarathy, 2017). Esse trabalho realizou um estudo empírico acerca de usuários potencialmente depressivos, procurando identificar se haveria diferença entre eles e um outro grupo de usuários, sendo esses não propiciados a terem um quadro de depressão. A pesquisa levou em consideração sinais como estilo linguístico, sinais emocionais, engajamento do usuário, geolocalização e topologia de rede para fazer a detecção de sintomas online de depressão. Então, puderam observar desvios significativos no comportamento de usuários deprimidos, que sinalizaram: padrões reduzidos e noturnos de atividade on-line, redução da participação ativa e passiva na rede, entre outros fatores.

(Mendes, 2021) descreve através de uma monografia, experimentos conduzidos para automatizar a classificação de postagens do *Twitter*, identificando conteúdos associados a sintomas de depressão, assim como a classificação de postagens de usuários depressivos. Os resultados indicaram que a regressão logística obteve os melhores desempenhos, alcançando uma média de F1 de 57% na classificação de sintomas e F1 de 64% na classificação de depressão. Ele oferece uma abordagem promissora para a identificação automatizada de conteúdo relacionado à depressão nas redes sociais, usando técnicas de aprendizado de máquina.

(Araujo *et al.*, 2012) apresentou um olhar amplo em relação ao tema de AS enquanto busca por potencialidades em aplicar esse tema na área de saúde. A pesquisa de revisão narrativa, selecionou 5 artigos nas bases de dados PubMed, ISI, ACM e IEEE, a qual relacionava a AS para diferentes temas. Os autores puderam observar que a mineração de opiniões aplicada aos fóruns que os usuários costumam participar, pode gerar mapeamentos de ocorrência de doenças podendo vir a ser utilizado na detecção de determinados perfis virtuais.

Em (Leiva, 2017) foi criado um modelo de classificador, que trazia como proposta a detecção precoce de usuários depressivos nas mídias sociais. O autor conferiu enfoque a plataforma *Reddit* utilizando diferentes algoritmos, como por exemplo o *Random Forest* (RF) e o *K-Nearest Neighbors*. O primeiro algoritmo apresentou o valor mais alto de precisão, e o segundo é aquele com maior *recall*.

3 Conceitos fundamentais

Dada a construção do conjunto de dados para o treinamento de um modelo de classificação, inicia-se a etapa de escolha do classificador. Para os experimentos foram utilizados os seguintes classificadores: MLP, RF, *Logistic regression* (LR), *Stochastic Gradient Descent* (SGD), e o NB.

A primeira técnica escolhida foi a do MLP que é uma arquitetura de rede neural artificial. A estrutura do MLP é composta por várias camadas de neurônios, organizadas em três tipos principais: a camada de entrada, que recebe os dados de entrada, que neste caso seriam os atributos da nossa base de dados sendo cada um representado por um neurônio na camada; a camada oculta, sendo a camada intermediária que recebe de entrada a saída da camada anterior, modificando assim esses dados por meio de pesos, realizando ajustes lineares e não lineares; e por fim a camada de saída que fornece os resultados, a quantidade de neurônios nessa camada vai depender da complexidade do problema (Haykin, 1999). Neste estudo, seguindo o artigo de Casani, examinou-se a eficácia dos modelos MLP e SVM na classificação de usuários depressivos. Foram empregados 10 subconjuntos no MLP, e a análise constatou consistentemente um desempenho superior a 90% de acurácia nos testes, reforçando a qualidade desses modelos.

Através do artigo de (Leiva, 2017), o algoritmo de aprendizado supervisionado *Decision Trees* (DT), cria um conjunto de múltiplas árvores de decisão. O princípio fundamental da RF é que a colaboração de inúmeros padrões independentes se sobrepõe ao desempenho de cada padrão isolado. Ou seja, os possíveis equívocos individuais de cada DT ocorrem isoladamente e ela como um todo não é influenciada por esse erro, pois o resultado final é direcionado de acordo com a maioria dos resultados individuais de cada modelo (Breiman, 2001), e por isso essa foi mais uma das técnicas escolhidas para estar sendo avaliada neste trabalho.

A LR, o qual é comumente aplicada em problemas de classificação, onde tem como objetivo prever a que categoria pertence determinada situação, é um método o qual é utilizado para lidar com relacionamentos entre uma variável categórica que é dependente e outras variáveis que são independentes. A partir de uma função logística que transforma a saída de um modelo linear em uma probabilidade, essas variáveis dependentes vão ter sua probabilidade de pertencer a uma determinada categoria estimada, tendo os valores previstos em um intervalo de 0 e 1 (Fernandes *et al.*, 2020).

No treinamento do modelo de LR tem-se o ajuste dos coeficientes do conjunto de dados de treinamento de forma iterativa, para diminuir a diferença entre as previsões do modelo e as categorias reais observadas nos dados. Por ter sido um modelo de rápido treinamento, essa técnica também foi colocada em avaliação neste trabalho.

O SGD é uma variação do *Gradient Descent* tradicional, mas em vez de calcular o gradiente usando todo o conjunto de dados (como ocorre no *Gradient Descent* convencional), o SGD usa apenas uma única amostra aleatória de dados por vez para atualizar os parâmetros do modelo. Isso torna o algoritmo rápido, possibilitando o uso em dados esparsos, e eficiente, especialmente para conjuntos de dados grandes, já que usa apenas uma amostra/instância em cada etapa para calcular o gradiente, o que o torna computacionalmente mais leve em comparação com métodos que usam o conjunto de dados completo a cada iteração (Melo, 2022).

A última técnica escolhida para a experimentação deste trabalho foi o *Naive Bayes*, que é um algoritmo de classificação probabilístico baseado no Teorema de

Bayes, que assume independência entre os recursos (ou atributos) utilizados para fazer a predição. A ideia fundamental por trás do NB é utilizar a probabilidade condicional para calcular a probabilidade de um ponto de dados pertencer a uma determinada classe com base nas evidências fornecidas pelos recursos (Johannben *et al.*, 2018).

4. Metodologia

Este capítulo apresenta a aplicação de PLN para detecção de PPD em mídias sociais. Na Seção 4.1 é descrita a proposta deste trabalho. A compreensão dos dados é apresentada na Seção 4.2. Na Seção 4.3 é detalhado as estratégias adotadas para limpar a base de dados, visando aprimorar o conjunto de dados, juntamente com o processo de *embedding* na Seção 4.4. Por fim, são apresentadas a configuração experimental e critérios de avaliação, nas Seções 4.5 e 4.6, respectivamente.

4.1 Proposta de pesquisa

Com a grande popularidade das mídias sociais, que trazem consigo uma grande gama de informações e assuntos, é inevitável o encontro de diferentes indivíduos que carregam perspectivas que podem vir a instigar fatores de risco para a saúde mental e o comportamento suicida (Vedana, 2018). Nessas mídias sociais, a extração de emoções tanto positivas quanto negativas pode ser feita através de textos com a utilização da AS, que coloca em prática técnicas de PLN para realizar o processamento e categorização dos dados.

De acordo com (Shron, 2014), a compreensão da essência do negócio em questão é de suma importância. Sem uma compreensão sólida das metas e desafios do domínio, a mineração de dados corre o risco de se tornar uma atividade sem direção, com potencial para desperdiçar esforços em análises irrelevantes ou resultados que não agregam valor aos objetivos organizacionais. Ao obter um entendimento abrangente das necessidades e metas do negócio, é possível direcionar corretamente o processo de mineração de dados para entregar resultados relevantes e valiosos.

Nesse sentido, este trabalho aplica técnicas de processamento de linguagem natural como: *Multilayer Perceptron*; *Random Forest*; LR; *Naive bayes*; e *Stochastic gradient descent* (SGD), visto que essas técnicas estão em enfoque neste tipo de linha de pesquisa, de acordo com os estudos referenciados neste trabalho.

O principal desafio enfrentado pela literatura no contexto deste trabalho está na obtenção de uma base de dados para treinamento. Isto porque como se trata de um assunto tão delicado quanto a identificação de PPD, a identificação dos mesmos em textos, sendo eles positivos ou negativos, se torna uma identificação rasa, onde os algoritmos de Aprendizagem de Máquina (AM) não aprendem de um texto o que podem ser sinais depressivos.

Com o intuito de fazer os algoritmos escolhidos para o experimento deste trabalho obterem um bom nível de aprendizado, e conseqüentemente, uma acurácia elevada em seus resultados, foi escolhido trabalhar com o conjunto de treinamento fornecido pelo projeto que disponibilizou a base de dados, onde os sintomas de usuários com PPD estão identificados nos dados textuais através de um especialista, que vem para garantir uma assertividade maior nos resultados e uma entrega de qualidade.

4.2 Entendimento dos dados

A base de dados coletada foi do projeto chamado *Sentiment140* (Go, A.; Bhayani, R.; e Huang, L., 2009), o qual utiliza classificadores construídos para algoritmos de AM, diferentes de outros produtos que usam de forma simplificada uma abordagem baseada em palavras-chave, que por sua vez possuem maior precisão, porém menor *recall*. Esse projeto fornece transparência para os resultados de classificação individuais dos *Tweets*, facilitando a avaliação da precisão dos classificadores.

O *Sentiment140* foi criado pelos cientistas da computação Alec Go, Richa Bhayani, e Lei Huang, o qual não deixaram como código aberto seu experimento, porém disponibilizaram suas bases de treino e teste.

A base de dados conta com 1.600.000 padrões e pode ser descrita como multivariada, do qual a problemática envolvida é de classificação, ou seja, pretende-se prever uma classe ou atributo categórico que determinados dados pertencem, que no caso deste trabalho seria identificar usuários com PPD.

O conjunto de dados de treino conta com 1.280.000 padrões, sendo 6 (seis) colunas, são elas: coluna de polaridade, que neste caso é o *target* (o qual classifica o *tweet* sendo 0 = negativo e 1 = positivo); coluna de ID; a coluna data; a coluna *query*; a coluna usuário; e a coluna texto, que armazena o conteúdo da postagem. Em contrapartida, o conjunto de testes conta com 320.000 padrões, sendo as mesmas seis colunas. Dentre estes, a seleção das colunas de maior peso e conseqüentemente a remoção das consideradas desnecessárias, e que possuem baixa relação com o problema que pretende-se trabalhar (que neste caso foram as colunas ID, data e usuário) foi feita.

4.3 Pré-Processamento

Dois conjuntos de dados em língua inglesa foram usados para realizar os experimentos. Para os estudos, ambos conjuntos de dados passaram por algumas etapas de preparação descritas nos próximos parágrafos.

Para dar continuidade ao estudo, antes de realizar-se qualquer construção de modelo e seu conseqüente treinamento, foi-se necessário realizar um pré-processamento textual, com alguns processos comuns de limpeza de dados. Esse processo incluía várias ações:

- Remoção de pontuação, caracteres especiais, URLs e hashtags;
- Eliminação de espaços ou tabulações extras nos lados esquerdo e direito;
- Correção de erros de digitação e gírias;
- Conversão de palavras abreviadas para suas formas completas.

Através do *Natural Language Toolkit* (NLTK), que é uma das ferramentas/bibliotecas mais populares em Python para PLN (Konopp, 2022), foi realizado um processo para remover, do vocabulário inglês, uma lista de palavras irrelevantes e genéricas, como por exemplo: *i, you, a/an, the, he, which* e dentre outros. Importante ressaltar que recursos como *Stemming* e lematização, que vão reduzir palavras fatiando o final ou o início das palavras com a intenção de remover afixos (i.e. *stem*), ou reduzir a sua forma base do dicionário (i.e. *lemma*), também foram utilizados nessa parte de limpeza.

Outro ponto é o balanceamento, que a partir dos dados originais avaliará se há a necessidade do uso de técnicas de *oversampling*, para a criação de padrões novos na

classe minoritária, ou *undersampling*, para diminuir a quantidade de padrões da classe majoritária, o que não foi preciso visto que ambas as bases encontravam-se balanceadas. Também foi considerada a viabilidade de reordenar os atributos para tornar as bases de dados mais claras.

Foram criadas duas colunas, sendo utilizadas para montar o gráfico da Figura 1, com o objetivo de conferir um entendimento e uma melhor análise sobre os dados. A primeira coluna chamada “*word_count*”, o qual se refere a uma coluna que vai armazenar a quantidade de palavras por *tweets* da base de dados, e a segunda coluna “*char_count*”, irá armazenar o número total de caracteres em um *tweet*.

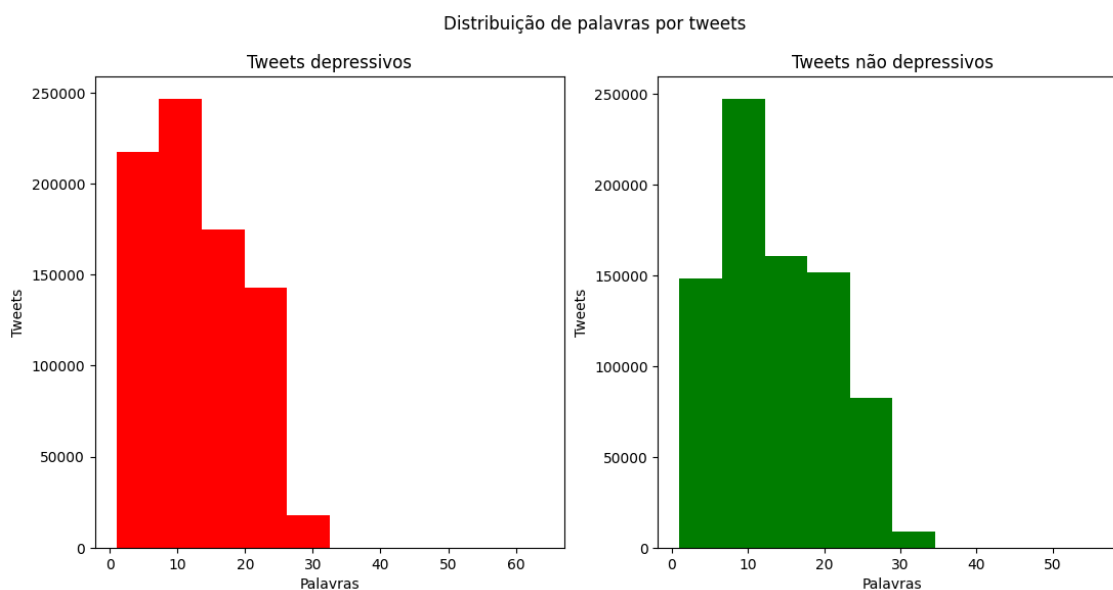


Figura 1: Gráfico de palavras por tweet
Fonte: O próprio autor.

4.4 *Embedding* dos dados

Para utilizarmos modelos estatísticos ou de *deep learning* em PLN, precisa-se de uma forma estruturada para se fazer o armazenamento de informações. Sendo os textos um tipo de dado não estruturado, realizou-se um processo de conversão dos dados textuais em dados/vetores numéricos bem definidos. Esse processo é conhecido como vetorização ou *Word Embedding*.

Com o objetivo de criar um conjunto de dados que pudesse ser utilizado no treinamento dos modelos deste trabalho, foi necessária a realização da etapa de pré-processamento de dados descrita na seção anterior. Esses dados também sofreram um processo de tokenização, onde cada uma das palavras dos *tweets* foram representadas como *tokens*, ou seja, uma *string* individual em uma lista.

No presente trabalho, os experimentos foram realizados com duas técnicas de *word embedding* e devidamente avaliadas, são elas: *Term Frequency-Inverse Document Frequencies* (TF-IDF) e *Word2Vec* (W2V).

O TF-IDF se refere a medidas estatísticas usadas para avaliar a relevância de uma palavra em um determinado documento, como neste caso, um *tweet*. Essas métricas combinam a frequência do termo no texto com a sua importância em relação a

todos os textos. Em resumo, essa abordagem concentra-se na frequência da palavra em todos os textos a serem analisados (Naeem *et al.*, 2022).

A segunda técnica é um algoritmo usado para representar palavras como vetores numéricos em um espaço vetorial de alta dimensão. O Word2Vec captura eficientemente relações semânticas e sintáticas entre palavras com base no contexto em que aparecem, isso acontece porque em seu treinamento o modelo ajusta os vetores de palavras de modo que palavras semanticamente e sintaticamente similares fiquem próximas umas das outras no espaço vetorial. Para esse tipo de técnica existem duas arquiteturas principais: *Continuous Bag of Words* (CBOW) e *Skip-gram*. O modelo CBOW tenta prever a palavra central com base no contexto das palavras vizinhas, já o *Skip-gram* tenta prever o contexto a partir de uma palavra central (Sivakumar *et al.*, 2020).

A arquitetura escolhida para o treinamento do W2V foi o *Skip-gram* por lidar melhor com *tokens* menos frequentes, capturando representações mais ricas para essas palavras. Além disso, o *Skip-gram* foi dado como o mais adequado para entender nuances linguísticas associadas à depressão, visto que o contexto e a relação entre as palavras têm um papel significativo na determinação do sentimento (Melo, 2022).

4.5 Configuração experimental

O experimento foi conduzido em duas etapas, realizando experimentos com o conjunto de dados utilizando duas técnicas de *word embedding*: TF-IDF e W2V. Em todos os experimentos, buscou-se aprimorar o desempenho dos modelos através da otimização de hiperparâmetros para encontrar as melhores combinações. Cada modelo foi avaliado usando validação cruzada, empregando o F1 *score* como critério principal para a comparação de desempenho.

Os valores dos hiperparâmetros testados para os classificadores foram selecionados considerando as peculiaridades de cada um, após a verificação das faixas utilizadas em outros trabalhos da área. Durante os experimentos, as mesmas faixas de valores dos parâmetros foram aplicadas às duas técnicas de *word embedding*, buscando a combinação de parâmetros que produz o maior valor de F1. As combinações de valores utilizadas em cada parâmetro na técnica TF-IDF e W2V estão apresentadas na Tabela 1.

Tabela 1. Hiperparâmetros utilizados nos classificadores para as técnicas TF-IDF e W2V

Classificadores	Identificador	Hiperparâmetros
<i>Logistic regression</i>	Penalty	['l1'; 'l2']
	C	[0.1; 1; 10]
	Solver	['liblinear'; 'saga'; 'lbfgs']
<i>Bernoulli Naive Bayes</i>	Alpha	[1; 0.3; 0,1]
	Class_count_	[10; 50; 100]
	Fit_prior	['True'; 'False']

<i>Random Forest</i>	n_estimators	[30; 50; 70]
	max_features	['auto'; 'sqrt'; 'log2']
	max_depth	[10; 50]
<i>Stochastic Gradient Descent</i>	Alpha	[1; 0.3; 0.1; 0.03; 0.003]
	Penalty	['l1'; 'l2'; 'elasticnet']
	eta0	[0.03; 0.01; 0.003; 0.001]
<i>Multilayer Perceptron</i>	Alpha	[0.0001; 0.05]
	Solver	['sgd'; 'adam']
	Activation	['tanh'; 'relu']
	Learning_rate	['constant'; 'adaptive']
	Hidden_layer_sizes	[(50,50,50); (50,100,50); (100,)]

O desenvolvimento de todas as etapas da proposta foi realizado na linguagem Python, utilizando as bibliotecas *SK-learn* (Géron, 2022) para os classificadores, e Gensim (Haider *et al.*, 2020) para geração dos *embeddings* dos dados. Os experimentos foram executados utilizando os serviços de uma máquina com a seguinte configuração: Intel core i5-9300 H, 2 núcleos de CPU, GPU Nvidia GeForce GTX 1650, 8GB de memória RAM e armazenamento do tipo SSD, com 225 GB de capacidade, e HD com 1 T de capacidade.

4.6 Critérios de avaliação

Com o objetivo de avaliar os experimentos, foi necessária a utilização de uma metodologia de avaliação que garantisse a validação dos experimentos com um mesmo tipo de método. A principal métrica de desempenho escolhida para os experimentos deste trabalho foi a métrica F1, que é baseada em outras duas métricas, são elas: precisão e *recall*. Ao combinar precisão e *recall*, o F1 favorece modelos que têm um desempenho equilibrado entre identificar corretamente as instâncias positivas e negativas, tornando-o uma métrica valiosa para situações em que tanto falsos positivos quanto falsos negativos são igualmente importantes.

A métrica Precisão (P), mede a proporção de exemplos positivos corretamente identificados pelo modelo em relação ao total de exemplos classificados como positivos pelo modelo, conforme a Equação 1, onde verdadeiros positivos (VP) e falsos positivos (FP). Em resumo, a precisão indica a precisão das predições positivas do modelo, medindo a qualidade das predições positivas em relação a todas as predições feitas.

A métrica *recall* (C), mede a proporção de exemplos positivos corretamente identificados pelo modelo em relação ao número total de exemplos que realmente pertencem à classe positiva, conforme mostrado na Equação 2, onde falsos negativos (FN). Em outras palavras, o *recall* indica a habilidade do modelo em identificar todos os exemplos relevantes da classe positiva.

A F1, pode ser vista como a média harmônica entre as métricas de precisão e *recall*, como mostrado na Equação 3. Uma outra métrica que vale a pena ser mencionada é a Acurácia, que é dada pela proporção de verdadeiros positivos em

relação ao total de previsões positivas, mostrada na Equação 4, onde verdadeiros negativos (VN). Abaixo é possível ver cada uma das equações para as presentes métricas.

$$\text{Precisão} = \text{VP} / (\text{VP} + \text{FP}) \quad (1)$$

$$\text{Recall} = \text{VP} / (\text{VP} + \text{FN}) \quad (2)$$

$$\text{F1} = 2 * ((\text{P} * \text{C}) / (\text{P} + \text{C})) \quad (3)$$

$$\text{Acurácia} = (\text{VP} + \text{VN}) / (\text{VP} + \text{VN} + \text{FP} + \text{FN}) \quad (4)$$

No presente trabalho a matriz de confusão foi analisada, permitindo-se uma análise mais ampla dos resultados obtidos no problema de classificação, onde se busca categorizar resultados em possibilidades limitadas existentes como: positivo ou negativo.

A validação cruzada (VC) foi escolhida para avaliar os modelos induzidos. Este método envolve a divisão da base de dados em k partes de dados de tamanhos iguais. Como resultado, o algoritmo é treinado usando $k - 1$ subconjuntos, e a parte restante é usada para validar o modelo inferido. O procedimento é repetido k vezes, alterando os subconjuntos de validação a cada iteração. O número de subconjuntos (k) utilizado para validar os modelos desenvolvidos para este trabalho é de 10 subconjuntos, valor frequentemente utilizado na literatura pertinente a este trabalho.

5. Experimentos e Resultados

Nesta seção, realizou-se a aplicação de duas técnicas de *embedding*, visando identificar a melhor para o problema proposto. Os classificadores passaram por uma etapa inicial de experimentação para se obter resultados com parâmetros padrão, e com base nos resultados, cada classificador selecionado foi submetido à otimização de hiperparâmetros, buscando alcançar o máximo desempenho para ambas as técnicas de *embedding*.

5.1 Resultados para a técnica de TF-IDF e W2V com hiperparâmetros padrão

Tabela 2. Resultados para a técnica W2V com hiperparâmetros padrão

Classificadores	Classe	Precisão	Recall	F1-Score	Acurácia
<i>Logistic regression</i> (Penalty = 12; C = 1; solver = lbfgs)	0	0,50	0,50	0,50	0,50
	1	0,50	0,50	0,50	
<i>Bernoulli Naive Bayes</i> (Alpha = 1; Class_count_ = 50; fit_prior = True)	0	0,50	0,51	0,50	0,50
	1	0,50	0,49	0,49	
<i>Random Forest</i> (n_estimators = 100; max_features =	0	0,50	0,52	0,51	0,50
	1	0,50	0,48	0,49	

sqrt; max_depth = None)					
<i>Stochastic Gradient Descent</i> (Alpha = 0.0001; Penalty = 12; eta0 = 0.0)	0	0,50	0,16	0,24	0,50
	1	0,50	0,84	0,63	
<i>Multilayer Perceptron</i> (Alpha = 0.0001; Solver = adam; Activation = relu; Learning_rate = constant; Hidden_layer_sizes = (100,))	0	0,50	0,43	0,46	0,50
	1	0,50	0,57	0,53	

Na Tabela 2 é possível ver os resultados de diferentes modelos de aprendizado de máquina utilizando a técnica de W2V com parâmetros padrão. Cada modelo é avaliado em termos de precisão, *recall*, *F1-Score* e acurácia para duas classes (0 e 1). Esses resultados sugerem um desempenho muito semelhante entre os modelos, todos com pontuações em torno de 0,50 para as métricas de interesse. Isso requer, no futuro, uma análise mais aprofundada para identificar possíveis problemas ou realizar novos ajustes nos modelos. Contudo, é válido ressaltar que o classificador SGD teve o maior valor para *recall* e *F1-Score* para classificar a classe 1, sendo 0,84 e 0,63 respectivamente.

Tabela 3. Resultados para a técnica TF-IDF com hiperparâmetros padrão

Classificadores	Classe	Precisão	<i>Recall</i>	<i>F1-Score</i>	Acurácia
<i>Logistic regression</i> (Penalty = 12; C = 1; solver = lbfgs)	0	0,78	0,76	0,77	0,77
	1	0,77	0,78	0,77	
<i>Bernoulli Naive Bayes</i> (Alpha = 1; Class_count_ = 50; fit_prior = True)	0	0,75	0,80	0,77	0,76
	1	0,78	0,73	0,76	
<i>Random Forest</i> (n_estimators = 100; max_features = sqrt; max_depth = None)	0	0,70	0,50	0,59	0,64
	1	0,61	0,79	0,69	
<i>Stochastic Gradient Descent</i> (Alpha = 0.0001; Penalty = 12; eta0 = 0.0)	0	0,78	0,58	0,66	0,71
	1	0,67	0,84	0,74	
<i>Multilayer Perceptron</i> (Alpha = 0.0001; Solver = adam; Activation = relu; Learning_rate = constant; Hidden_layer_sizes = (100,))	0	0,77	0,74	0,75	0,76
	1	0,75	0,78	0,76	

Na Tabela 3 é possível ver os resultados de diferentes modelos de aprendizado de máquina utilizando a técnica TF-IDF com parâmetros padrão. Cada modelo é avaliado em termos de precisão, *recall*, *F1-Score* e acurácia para duas classes (0 e 1). Esses resultados mostram que os modelos possuem desempenho aproximadamente semelhantes para algumas métricas, porém os classificadores LR e SGD apresentaram o valor 0,78 como o maior valor de *precisão* para a classe 0 em comparação com os outros classificadores. Além disso, o SGD se mostrou o classificador com melhor resultado para classificar a classe 1, com um valor de 0,84. Contudo, o LR e o *Bernoulli Naive*

Bayes foram os modelos com *F1-Score* mais altos para o conjunto de dados utilizados neste trabalho, marcando um valor de 0,77.

Esse conjunto de resultados, no qual TF-IDF e W2V foram avaliados com modelos utilizando configuração padrão de hiperparâmetros, mostrou que o TF-IDF obteve resultados consistentemente mais altos, apresentando um ganho de percentual 27% superior em relação ao resultado do W2V. Na próxima subseção serão exibidos os resultados das técnicas de *embedding* combinadas com modelos que tiveram seus hiperparâmetros ajustados.

5.2 Resultados para a técnica de TF-IDF e W2V com os melhores hiperparâmetros

Tabela 4. Resultados para a técnica W2V com otimização de hiperparâmetros

Classificadores	Classe	Precisão	Recall	F1-Score	Acurácia
<i>Logistic regression</i> (Penalty = l2; C = 1; solver = liblinear)	0	0,50	0,50	0,50	0,50
	1	0,50	0,50	0,50	
<i>Bernoulli Naive Bayes</i> (Alpha = 0,3; Class_count_ = 50; fit_prior = False)	0	0,50	0,50	0,50	0,50
	1	0,50	0,50	0,50	
<i>Random Forest</i> (n_estimators = 50; max_features = sqrt; max_depth = 10)	0	0,50	0,51	0,50	0,50
	1	0,50	0,49	0,50	
<i>Stochastic Gradient Descent</i> (Alpha = 0,003; Penalty = l2; eta0 = 0,03)	0	0,50	0,06	0,10	0,50
	1	0,50	0,95	0,65	
<i>Multilayer Perceptron</i> (Alpha = 0,0001; Solver = sgd; Activation = tanh; Learning_rate = constant; Hidden_layer_sizes = (50,50,50))	0	0,50	0,22	0,30	0,50
	1	0,50	0,78	0,61	

Na Tabela 4 é possível ver os resultados de diferentes modelos de aprendizado de máquina utilizando a técnica de W2V com hiperparâmetros otimizados. Cada modelo é avaliado em termos de precisão, *recall*, *F1-Score* e acurácia para duas classes (0 e 1). Estes resultados, mesmo com a otimização de hiperparâmetros, mostram um desempenho uniformemente baixo para todos os modelos, tendo a média geral de F1 de todos os modelos treinados em 0,50. Isso indica que os ajustes nos hiperparâmetros não foram suficientes para melhorar significativamente o desempenho dos modelos de W2V neste conjunto de dados.

Dando destaque para o classificador SGD, que apresentou 0,95 para a métrica *recall* e 0,65 para o *F1-Score*, sendo esses os maiores valores encontrados no treinamento com a otimização de hiperparâmetros para essa técnica. É possível que a natureza dos dados ou a escolha dos modelos não seja ideal para a técnica de W2V, indicando que outros métodos de representação vetorial ou modelos de aprendizado de máquina podem ser mais apropriados para este problema específico.

Tabela 5. Resultados para a técnica TF-IDF com otimização de hiperparâmetros

Classificadores	Classe	Precisão	Recall	F1-Score	Acurácia
<i>Logistic regression</i> (Penalty = l2; C = 1; solver = liblinear)	0	0,79	0,80	0,79	0,79
	1	0,77	0,79	0,78	
<i>Bernoulli Naive Bayes</i> (Alpha = 0,3; Class_count_ = 50; fit_prior = False)	0	0,75	0,79	0,77	0,77
	1	0,78	0,74	0,76	
<i>Random Forest</i> (n_estimators = 50; max_features = sqrt; max_depth = 10)	0	0,68	0,64	0,66	0,67
	1	0,66	0,70	0,68	
<i>Stochastic Gradient Descent</i> (Alpha = 0,003; Penalty = l2; eta0 = 0,03)	0	0,74	0,68	0,71	0,72
	1	0,70	0,77	0,73	
<i>Multilayer Perceptron</i> (Alpha = 0,0001; Solver = sgd; Activation = tanh; Learning_rate = constant; Hidden_layer_sizes = (50,50,50))	0	0,77	0,79	0,78	0,79
	1	0,79	0,77	0,78	

Na Tabela 5 é possível ver os resultados de diferentes modelos de aprendizado de máquina utilizando a técnica TF-IDF com hiperparâmetros otimizados. Cada modelo é avaliado em termos de precisão, *recall*, *F1-Score* e acurácia para duas classes (0 e 1). Em geral, os modelos tiveram um desempenho razoável, com valores de acurácia variando entre 0.67 e 0.79. Isso sugere que os modelos estão conseguindo classificar as instâncias corretamente.

Há uma variação considerável no desempenho entre os diferentes modelos. Por exemplo, o Random Forest teve um desempenho relativamente baixo em comparação com outros modelos, com acurácia de apenas 0.67. Enquanto isso, o Multilayer Perceptron e o LR obtiveram a maior acurácia geral de 0.79.

Comparativamente aos resultados anteriores com TF-IDF e W2V, os resultados após a otimização de hiperparâmetros mostram melhorias consideráveis no desempenho dos modelos de TF-IDF, apresentando um ganho percentual de 2% em relação ao resultado dos experimentos com hiperparâmetros padrão. Esses resultados sinalizam a importância do ajuste de hiperparâmetros na obtenção de modelos mais precisos e eficazes para a tarefa de classificação. A seguir estão listadas as principais melhorias observadas a partir da otimização;

- Melhorias no Desempenho:
 - A precisão, recall e F1-Score melhoraram significativamente em comparação com os resultados anteriores, indicando que a otimização de hiperparâmetros foi eficaz para melhorar o desempenho dos modelos.
 - Em geral, os modelos apresentaram uma capacidade melhor de discriminar entre as classes 0 e 1, resultando em métricas de avaliação mais altas.

- Variação de Desempenho entre Modelos:
 - Alguns modelos, como Logistic Regression e Multilayer Perceptron, demonstraram um desempenho mais consistente e melhor em todas as métricas avaliadas em comparação com Naive Bayes, Random Forest e Stochastic Gradient Descent, tendo a mesma acurácia de 0,79 para ambos. Entretanto, o LR apresentou 0,79, 0,80, e 0,79 para precisão, recall, e F1-Score respectivamente, como os valores mais altos encontrados no treinamento para a classe 0, quando comparados com os outros modelos.

6. Conclusão

A depressão muitas vezes é subdiagnosticada e subtratada devido a estigmas sociais e à falta de detecção precoce por profissionais de saúde. Este trabalho buscou aplicar diferentes tipos de algoritmos inteligentes e dois métodos de pré-processamento textual, a fim de fortalecer a compreensão de como o domínio de PLN e AS podem dar apoio nesta área de estudo, identificando postagens de usuários com PPD.

Considerando os resultados alcançados, o classificador LR e MLP se destacam em comparação com o RF, SGD e NB para o conjunto de dados analisados. Onde o LR não apenas demonstrou um desempenho consistente para todas as técnicas avaliadas, alcançando valores de F1 similares ao MLP, mas também se apresentou como o modelo com os melhores valores. Essa eficiência no treinamento, combinada com resultados satisfatórios, posiciona o LR como uma escolha pragmática e eficaz para essa tarefa.

Este trabalho enfrentou algumas limitações devido a diversidade e a complexidade do comportamento humano, especialmente em relação à depressão, que é um assunto que pode levar a nuances que os algoritmos não capturam totalmente, pois nem sempre os sintomas são expressos de maneira direta ou óbvia nas mídias sociais. Apesar dos resultados obtidos, realizar aplicações similares em diferentes áreas com problemas parecidos, representa uma direção promissora para trabalhos futuros, especialmente no contexto da Inteligência Artificial Explicável (XAI, do inglês, *Explainable Artificial Intelligence*). A inclusão de XAI em tais aplicações pode permitir compreender detalhes sobre o funcionamento interno dos classificadores utilizados.

A análise da quantidade de palavras por *tweet* da base de dados pode ser um outro ponto a ser explorado, visto que se observou certo padrão em *tweets* de caráter depressivo e não depressivo, quanto a sua quantidade de caracteres e palavras. Além disso, a análise e comparação de classificadores mais complexos, por exemplo utilizando Aprendizagem Profunda, é outra área de pesquisa sugerida, assim como explorar a utilização de outras técnicas de pré-processamento textual. Os trabalhos futuros sugeridos podem vir a contribuir significativamente para novas aplicações de AS e PLN na detecção de indícios de depressão a partir de textos.

Referências

ARAUJO, G. D. *et al.* Análise de sentimentos sobre temas de saúde em mídia social. **Journal of Health Informatics**, v. 4, n. 3, 2012.

BOUKIL, Samir et al. Deep learning algorithm for suicide sentiment prediction. In: **Advanced Intelligent Systems for Sustainable Development (AI2SD'2018) Vol 4:**

Instituto Federal de Educação, Ciências e Tecnologia de Pernambuco. Campus Paulista. Curso de Análise e Desenvolvimento de Sistemas. 22 de dezembro de 2023. 15

Advanced Intelligent Systems Applied to Health. Springer International Publishing, 2019. p. 261-272.

BREIMAN, L. Random forests. **Machine Learning**, Kluwer Academic Publishers, Hingham, MA, USA, v. 45, n. 1, p. 5–32, out. 2001.

CARMO NETO, Francisco Vieira do. Análise de sentimentos de reviews de produtos de e-commerces brasileiros. 2022.

CASANI, Vinicius; *et al.* DP-Symptom-Identifier: Uma estratégia para classificar sintomas de depressão utilizando um conjunto de dados textuais na língua portuguesa. In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA. Porto Alegre: Sociedade Brasileira de Computação, (2021), p. 153-161.

CHOUDHURY, Munmun de; COUNTS, Scott; HORVITZ, Eric. Social media as a measurement tool of depression in populations. In: **Proceedings of the 5th annual ACM web science conference**. 2013. p. 47-56.

DUPAS, Gilberto. **Ética e poder na sociedade da informação**: de como a autonomia das novas tecnologias obriga a rever o mito do progresso. 2. ed. rev. e ampl. São Paulo: UNESP, 2001.

FERNANDES, A. A. Tôrres; FILHO, D. B. Figueiredo; DA ROCHA, E. Carvalho; NASCIMENTO, W. da Silva. Read this paper if you want to learn logistic regression. **Revista de Sociologia e Política**, v. 28, n. 74, 2020.

GÉRON, Aurélien. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**. " O'Reilly Media, Inc.", 2022.

GO, A.; BHAYANI, R.; e HUANG, L., 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford, 1(2009), p.12*.

HAIDER, Mofiz Mojib *et al.* Automatic text summarization using gensim word2vec and k-means clustering algorithm. In: **2020 IEEE Region 10 Symposium (TENSYP)**. IEEE, 2020. p. 283-286.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. [S.l.: s.n.], 1999.

JOHANNßEN, Dirk; BIEMANN, Chris. Between the lines: Machine learning for prediction of psychological traits—a survey. In: **Machine Learning and Knowledge Extraction: Second IFIP TC 5, TC 8/WG 8.4, 8.9, TC 12/WG 12.9 International Cross-Domain Conference, CD-MAKE 2018, Hamburg, Germany, August 27–30, 2018, Proceedings 2**. Springer International Publishing, 2018. p. 192-211.

JÚNIOR, M. A. G. N. *et al.* Depressão em estudantes de medicina. **Rev Med Minas Gerais**, v. 25, n. 4, p. 562-567, 2015.

KONOPP, Jhonatan da Silva *et al.* Aplicação da Análise de Sentimento para Avaliar Mensagens Significativas em um Ambiente Colaborativo: um estudo de caso no ambiente Collabora. **RENOTE**, v. 20, n. 1, p. 132-142, 2022.

LEIVA, V.; FREIRE, A. Towards suicide prevention: Early detection of depression on social media. In: Internet Science. Cham: Springer International Publishing, 2017. p. 428–436. ISBN 978-3-319-70284-1.

LÉVY, Pierre. **O que é o virtual?**. 2. ed. São Paulo: Editora 34, 2011. ISBN 978-8573260366.

LIU, B. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012.

MCGIRR, Alexander; TURECKI, Gustavo. The relationship of impulsive aggressiveness to suicidality and other depression-linked behaviors. **Current psychiatry reports**, v. 9, n. 6, p. 460-466, 2007.

MELO, Ricardo Luiz da Silva. **UMA ABORDAGEM PARA EXTRAÇÃO DE RELAÇÕES ENTRE ENTIDADES NOMEADAS UTILIZANDO AUTOENCODER E GRADIENT BOOSTING**. 2022. Tese de Doutorado. Universidade Federal do Rio de Janeiro.

MENDES, Augusto Rozendo. Detecção automática de postagens possivelmente depressivas em redes sociais. 2021

NAEEM, Muhammad Zaid *et al.* Classification of movie reviews using term frequency-inverse document frequency and optimized machine learning algorithms. **PeerJ Computer Science**, v. 8, p. e914, 2022.

PIGA, Bruna Maria Fava; SHIMA, Vivian Taciany Bonassoli; ROMANICHEN, F. M. D. F. Análise das prescrições de ansiolíticos e antidepressivos antes e durante a pandemia da COVID-19 Analysis of prescriptions for anxiolytics and antidepressants before and during the COVID-19 Pandemic. **Brazilian Journal of Development**, v. 7, n. 11, p. 107178-107193, 2021.

QUEMEL, Gleicy Kelly China *et al.* Revisão integrativa da literatura sobre o aumento no consumo de psicotrópicos em transtornos mentais como a depressão. **Brazilian Applied Science Review**, v. 5, n. 3, p. 1384-1403, 2021.

SHRON, Max. **Thinking with Data: How to Turn Information into Insights**. " O'Reilly Media, Inc.", 2014.

SILVA, Aldilene Clemente da *et al.* Uso de dispositivos eletrônicos & redes sociais digitais: possíveis efeitos positivos e colaterais. 2021.

SIVAKUMAR, Soubraylu *et al.* Review on word2vec word embedding neural net. In: **2020 international conference on smart electronics and communication (ICOSEC)**. IEEE, 2020. p. 282-290.

VEDANA, Kelly Graziani Giacchero. Mídias sociais e suicídio: Universidade de São Paulo. Escola de Enfermagem de Ribeirão Preto. **SMAD, Revista Eletrônica Saúde Mental Álcool e Drogas (Edição em Português)**, v. 14, n. 4, p. 194-194, 2018.

VEDULA, N.; PARTHASARATHY, S. Emotional and linguistic cues of depression from social media. In: DH. [S.l.: s.n.], 2017.