

Identificação de Perfis Depressivos em Redes Sociais Utilizando Aprendizado de Máquina: um Mapeamento Sistemático

Vinicius Casani

Universidade Tecnológica Federal do Paraná
Dois Vizinhos, PR, Brasil
viniciuscasani@alunos.utfpr.edu.br

Alinne Cristinne Correa Souza

Universidade Tecnológica Federal do Paraná
Dois Vizinhos, PR, Brasil
alinnesouza@utfpr.edu.br

Rafael Gomes Mantovani

Universidade Tecnológica Federal do Paraná
Apucarana, PR, Brasil
rafaelmantovani@utfpr.edu.br

Francisco Carlos Monteiro Souza

Universidade Tecnológica Federal do Paraná
Dois Vizinhos, PR, Brasil
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. In the literature, different techniques have been studying to analyze and recognize this disease such as Natural Language Processing, Sentiment Analysis, and Machine Learning. In this paper, we describe a systematic mapping to identify evidence regarding techniques that are often used to identify depressive profiles. We analyzed 472 studies and we selected 25 primary studies. These studies indicate that the SVM and NB techniques have been most used to detect possible depressive profiles in social networks. Furthermore, Twitter and Facebook with 35,5% and 22,6%, respectively were the social media most have been used by users' express their feelings regarding the most varied subjects.

KEYWORDS

Systematic Mapping, Sentiment Analysis, Natural Language Processing

1 INTRODUÇÃO

Considerada o Mal do Século pela Organização Mundial da Saúde (OMS), a depressão é um transtorno psiquiátrico que afeta o emocional da pessoa, onde a mesma passa a apresentar desânimo, baixa autoestima, impossibilidade de sentir alegria, sentimento de inutilidade ou culpa, e em casos graves, desejo de morrer, e que modifica a maneira como usam e processam a linguagem [1].

Pelo fato de existir essa modificação referente a linguagem, é possível utilizar a Análise de Sentimentos (AS) para identificar perfis depressivos em dados textuais, principalmente quando estes são extraídos de redes sociais, já que nelas diariamente são gerados uma quantidade exorbitante de dados e os usuários possuem uma maior facilidade para se expressar sobre os mais variados assuntos.

O processo de identificação dos perfis depressivos é realizado com o auxílio de técnicas de Processamento de Linguagem Natural (PLN) e de Aprendizado de Máquina (AM). A respeito deste interesse, foi realizado um Mapeamento Sistemático (MS), visando identificar quais técnicas de PLN e AM são frequentemente utilizadas para identificação de perfis depressivos. Neste trabalho são apresentados os resultados de um MS, os elementos essenciais do protocolo de pesquisa juntamente com a forma de conduzir o processo como um todo.

O presente artigo está organizado da seguinte forma: na Seção 2 são descritos os principais elementos do protocolo de pesquisa e como o MS foi conduzido. Na Seção 3 são discutidos os resultados alcançados e as eventuais ameaças à validade. Por fim as considerações finais são apresentadas nas Seção 4.

2 MAPEAMENTO SISTEMÁTICO

O MS visa apresentar identificar, avaliar e interpretar, de uma forma geral, evidências de uma área de pesquisa para um determinado tema, área ou fenômeno de interesse [2]. O objetivo deste MS é identificar as principais evidências disponíveis na literatura referentes à classificação de palavras, sentenças ou textos depressivos utilizando AS. Para auxiliar a condução deste MS foi utilizado o processo proposto por Kitchenham et. al [3], sendo composto por três fases: (i) Planejamento; (ii) Condução; e (iii) Análise.

2.1 Planejamento

Esta fase do MS consiste na formulação das questões de pesquisa e a definição da estratégia de busca, dos critérios de inclusão e exclusão. Para a condução deste MS as seguintes Questões de Pesquisa (QPs) foram desenvolvidas:

- **QP₁: Quais técnicas têm sido utilizadas para análise de sentimentos e identificação de perfis depressivos em redes sociais?**
- **QP₂: Quais tipos de rotulação de base de dados têm sido utilizadas para análise de sentimentos e identificação de perfis depressivos em redes sociais?**

Com a QP₁ espera-se identificar diferentes tipos de técnicas relacionadas à AS que têm sido utilizadas para identificar perfis depressivos em textos em redes sociais. A QP₂ visa identificar de onde são extraídas as informações textuais para criação das bases de dados utilizadas para para análise de sentimentos e identificação de perfis depressivos, e como estas bases estão sendo rotuladas, por exemplo rotulação manual, rotulação automática ou outros tipos.

A partir das QPs, foi gerada a *string* de busca que consistiu na combinação das palavras “*sentiment analysis*” e “*depression*” e seus respectivos sinônimos, todas no idioma inglês. Portanto, a seguinte *string* de busca foi gerada: (“*sentiment analysis*” OR “*data analysis*” OR “*data mining*”) AND (*depression* OR “*depressive mood*”).

Com a *string* definida, foi realizada a busca automática em quatro bases de dados eletrônicas (*ACM Digital Library*, *IEEE Xplore*, *Scopus* e *Springer*) de acordo com as diretrizes propostas por [4]. É importante destacar que nesta etapa foram selecionadas apenas quatro bases de busca, devido a quantidade de tempo limitada para a elaboração do mesmo.

Para auxiliar o processo de seleção de estudos relevantes, foram definidos os seguintes Critérios de Inclusão e Exclusão:

- **Critérios de Inclusão (CI):**
 - **CI₁:** estudos primários que apresentem técnicas para auxiliar a análise de sentimentos e a identificação de perfis depressivos em redes sociais;
 - **CI₂:** estudos primários que apresente pelo menos uma forma de categorizar sentimentos a partir de dados textuais.
- **Critérios de Exclusão (CE)**
 - **CE₁:** estudos primários que não sejam *full paper* ou *short paper* (pôsteres, tutoriais, dissertações e teses);
 - **CE₂:** estudos primários que não estejam escritos em inglês ou português;
 - **CE₃:** estudos primários que não estejam disponíveis; e
 - **CE₄:** estudos primários que seja uma versão anterior de um estudo mais completo sobre a mesma investigação.

2.2 Condução

O processo de condução do MS apresentado na Figura 1 consistiu na execução da *string* de busca nas bases selecionadas. É importante ressaltar que as buscas foram conduzidas no período de março até setembro de 2019.

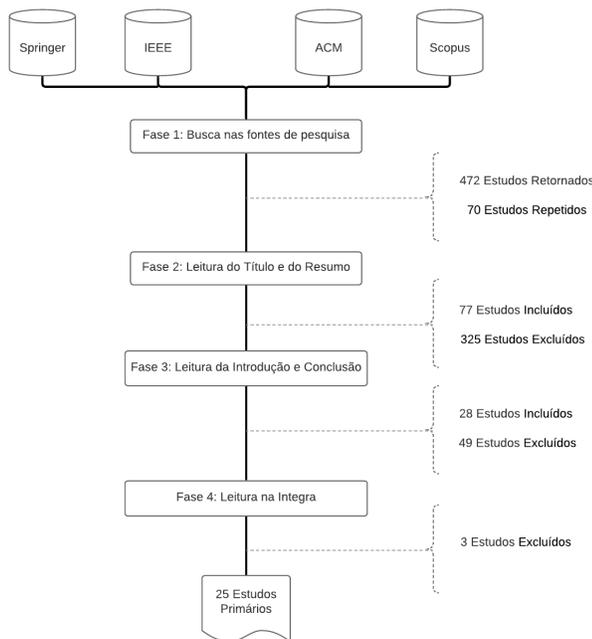


Figure 1: Processo de seleção dos estudos primários

Na fase 1 foram retornados 124 estudos na *IEEE*, 31 estudos na *ACM*, 88 estudos na *Springer* e 229 no *Scopus*. A partir dos 472 estudos retornados das fontes de busca, os estudos repetidos foram identificados e excluídos, os títulos e resumos dos estudos restantes (402) foram lidos e neles aplicados os critérios de inclusão e exclusão (Fase 2). Na Fase 3, foi realizada a leitura da introdução e conclusão dos 77 estudos que foram incluídos durante a fase 2 e neles foram aplicados os critérios de exclusão e inclusão. Na fase 4, os 28 estudos selecionados na etapa anterior foram lidos na íntegra, acarretando na exclusão de 3 estudos, totalizando 25 estudos primários incluídos.

3 RESULTADOS E DISCUSSÕES

Esta seção visa apresentar e discutir os resultados alcançados por meio do processo de classificação dos estudos. De acordo com os resultados, é notável que a maior concentração dos estudos incluídos foi por meio da base *IEEE* com 44% (11/25), seguida da *Scopus* com 32% (8/25), *Springer* com 16% (4/25) e *ACM* com 8% (2/25). Outra informação extraída foi o percentual de estudos primários selecionados de acordo com o tipo de publicação. 60% (15/25) dos estudos selecionados foram publicados em , conferências enquanto que 40% (10/25) foram em periódicos.

A partir dos 25 estudos primários analisados é possível observar que o primeiro estudo foi publicado no ano de 2013, a partir deste pode-se observar uma crescente em relação aos estudos publicados em cada ano, este efeito pode estar atribuído a popularização do tema, bem como uma maior preocupação da população com relação a depressão e suas consequências na sociedade global. O ano de 2017 é o que mais se destaca com 6 publicações realizadas. Nos anos de 2018 e 2019 houve uma queda no número de estudos publicados, conforme pode ser visto na Figura 2.

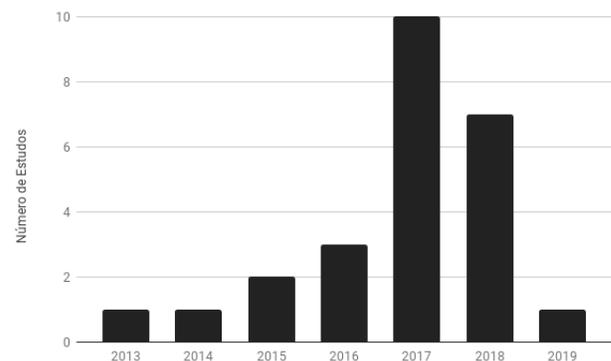


Figure 2: Estudos primários por ano de publicação

Na Figura 3 são apresentados o número de estudos por países. Para classificar o país do estudo, foi utilizado o país de origem do primeiro autor do estudo. A Austrália, por exemplo, se destaca por conter cinco estudos publicados, seguida pelo Estados Unidos com três estudos. Bangladesh, Coreia do Sul, Marrocos e Taiwan possuem duas publicações cada. Os demais países contem apenas um estudo publicado. Também pode-se observar que o Brasil possui apenas um estudo publicado, reiterando que pouco tem sido investigado nesta área no cenário brasileiro utilizando bases de dados na

língua portuguesa. Dentre os autores dos estudos, apenas Nguyen et. al possuem mais de uma publicação, a primeira em 2014 [5] e a segunda em 2017 [6].

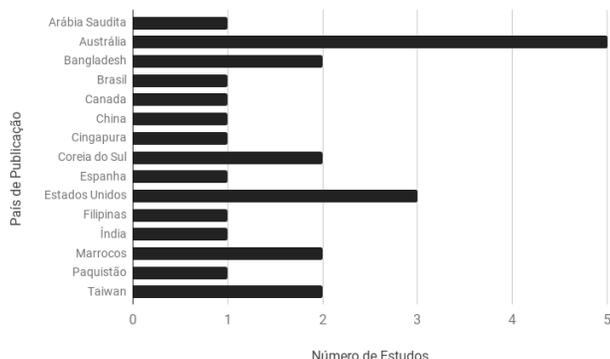


Figure 3: Estudos primários por país de publicação

Os 25 estudos selecionados são apresentados na Tabela 1, onde na primeira coluna tem-se a referência do estudo. Na segunda coluna são apresentada(s) a(s) técnica(s) utilizadas pelos estudos primários. Na terceira coluna são relatadas as base de dados, nas quais as postagens foram realizadas. Na quarta coluna são descritos os tipos de rotulação. Por fim, na coluna 5 tem-se a acurácia alcançada por cada estudo.

3.1 QP₁: Quais técnicas têm sido utilizadas para análise de sentimentos e identificação de perfis depressivos em redes sociais?

Para responder a QP₁, na Figura 4 são sumarizadas 39 técnicas diferentes utilizadas para AS no contexto deste trabalho. Dentre tais técnicas é importante destacar o algoritmo de classificação *Support Vector Machine (SVM)* que foi utilizado em 11 estudos, seguido pelo algoritmo *Naive Bayes (NB)*, utilizado em 9 estudos. Outros algoritmos que se destacam são o algoritmo *K Nearest Neighbor (KNN)* que aparece sendo utilizado por 5 estudos e o de *Decision Trees*, utilizado por 4 estudos.

Ainda pode-se observar que diversos estudos utilizaram técnicas de PLN para realizar o pré-processamento dos dados de extraídos das fontes de dados, como remoção de *stop-words*, *stemming* e entre outras. Mas alguns estudos como em [15], [17] e [26] utilizaram dicionários léxicos como principal forma de identificar os sentimentos nos textos, ou ainda com o objetivo de auxiliar na vetorização dos dados visando facilitar o processamento realizado por técnicas de AM.

3.2 QP₂: Quais são os tipos de rotulação das bases de dados que têm sido utilizadas?

Para a QP₂, é importante destacar que em alguns casos, o tipo de rotulação escolhida pelos autores dos estudos estava relacionada às fontes de dados utilizadas para extração das informações. A seguir é realizada uma análise dos dados identificados, bem como uma

breve descrição das fontes de dados. Por fim, são apresentadas as formas de rotulação utilizadas por cada um dos estudos.

Com base na análise realizada e nas informações apresentadas na Tabela 1, observa-se que postagens feitas no *Twitter* são utilizadas para classificação de depressão em 35,5% (11/25) dos estudos. Postagens realizadas no *Facebook* são utilizadas em 22,6% (7/25) dos estudos.

A comunidade *LiveJournal* é utilizada em 12,9% (4/25) dos estudos. Por outro lado, as postagens realizadas no *Reddit* compõem a base de dados de 9,7% (3/25) dos estudos analisados. Por fim, postagens realizadas no Sina Micro-Blog, no *blog PTT*, *blogs* de depressão em geral, o *dataset 20newsgroups*, coleções de notas em árabe e informações de vocabulários *online* são utilizadas em 3,2% (1/25) dos estudos.

A utilização do *Twitter* e *Facebook* em 58,1% (18/25) dos estudos, pode estar ligada ao fato de que ambas as redes sociais, possuem APIs públicas para coleta de dados, facilitando assim a construção das bases de dados tanto de treinamento como as de testes.

Postagens do *LiveJournal* também são frequentemente utilizadas, pois o *blog* possui um conjunto predefinido de 132 humores disponíveis para os usuários marcarem em suas postagens, auxiliando assim na classificação das bases. *LiveJournal* é uma comunidade virtual onde os usuários da Internet podem manter um blog, um jornal ou um diário. Cada usuário tem um perfil com o qual pode interagir com comunidades e outros usuários.

Reddit é uma comunidade de fóruns no qual os usuários podem criar tópicos para outros usuários interagirem. Sina Micro-Blog é um micro *blog* similar ao *Twitter* existente na China. PTT é um fórum de discussões chinês. Por fim, o *Center for Epidemiologic Studies Depression Scale (CES-D)* é um questionário de auto-relatos, que tem como finalidade medir a gravidade de sintomas depressivos na população em geral.

A análise referente as bases de dados citadas pelos estudos possibilitou identificar que apenas o estudo [13] teve sua base criada em português. Isto indica o *deficit* de estudos focados em analisar sentimentos no contexto de doenças mentais em português. Isso pode estar ligado ao fato da complexidade de realizar análises neste idioma, principalmente em publicações de redes sociais, já que o número de publicações com ambiguidade, figuras de linguagem, abreviações, entre outros aspectos é elevado neste contexto.

Para responder a QP₂, na Figura 5 são sumarizadas os dois tipos de rotulação das bases de dados. É importante destacar que somente dois estudos ([7] e [14]) não informaram como realizaram a rotulação das bases de dados. Os demais estudos foram divididos nas seguintes categorias: Rotulação Manual e Rotulação Automática, e são descritos a seguir.

• Rotulação Manual:

O estudo de Islam et. al [9] rotulou a base de dados com o auxílio de dois especialistas, que rotularam a base entre comentários que indicavam depressão e que não indicavam depressão, ao final, 58% dos comentários indicavam depressão e 42% não indicavam.

No estudo de Wang, Zhang e Sun [10] utilizou postagens de 50 usuários depressivos, e 50 de usuários não depressivos. Os usuários foram extraídos do conjunto de 180 usuários (90 depressivos e 90 não depressivos) que foram detectados por um grupo de psicólogos em [30].

Table 1: Visão geral dos estudos primários selecionados

Estudo	Técnica	Base Dados	Rotulação	Acurácia
[7]	SVM; NB; ME	Twitter; 20newsgroups	Não informado	SVM: 91%; NB: 83%; ME: 80%
[8]	NB; SVM	Twitter	Automática	NB: 83%; SVM: 79%
[9]	KNN	Facebook	Manual	60 - 70%
[10]	Linear Regression	Sina Micro-Blog	Manual	95%
[11]	Latent Dirichlet Allocation	Reddit	Manual	-
[12]	SVM; NB	Twitter; LiveJournal; Facebook	Manual	SVM: 67%; NB: 63%
[5]	Lasso	LiveJournal	Automática	-
[13]	NB;SVM;Random Forest; CNN-BLSTM-RNN - Softmax; CNN-BLSTM-RNN - SVM	Facebook	Automática	NB: 69%; SVM: 84%; Randon Forest: 81%; CNN BLSTM-RNN - Softmax: 89%; CNN BLSTM-RNN - SVM: 87%
[14]	NB; SVM; Lasso; Bayes Net; Decision Trees; KNN; Regression Analysis; Binary Logistic Regression; Rocchio Classifier; Perceptron classifier; Bagging; MANCOVA; K-means Algorithm; Latent Semantic Analysis	Facebook	Não informado	-
[15]	PCA juntamente com ANEW e LIWC	Twitter	Automática	-
[16]	Hierarchical Dirichlet Process	LiveJournal	Automática	-
[17]	Logistic Regression juntamente com Emolex/IsaWika Lexicon; LDA	Twitter; Facebook	Automática	99.92%
[18]	Decision Trees	Twitter	Automática	90%
[19]	SVM; KNN; Decision Trees	Facebook	Automática	60% a 80%
[20]	RNN	Facebook e CES-D	Automática	-
[21]	SVM; NB; Decision Trees	Twitter	Automática	90%
[6]	Lasso; NB; SVM; Logistic Regression	LiveJournal	Automática	-
[22]	PLN	Twitter; Vocabulários online	Manual	-
[23]	NB; SVM; Logistic Regression; KNN	Reddit	Manual	-
[24]	IBK, J48, SMO; PART	Twitter	Automática	-
[25]	PLN; SVM; Random Forest	Reddit	Manual	-
[26]	PLN juntamente com Lexicos	Fórum de discussão PTT	Manual	-
[27]	CNN; NB; KNN	Notas escritas em Árabe	Manual	-
[28]	PLN juntamente com LDA	Twitter e blogs de depressão	Automática	-
[29]	SVM; LIWC	Twitter	Manual	-

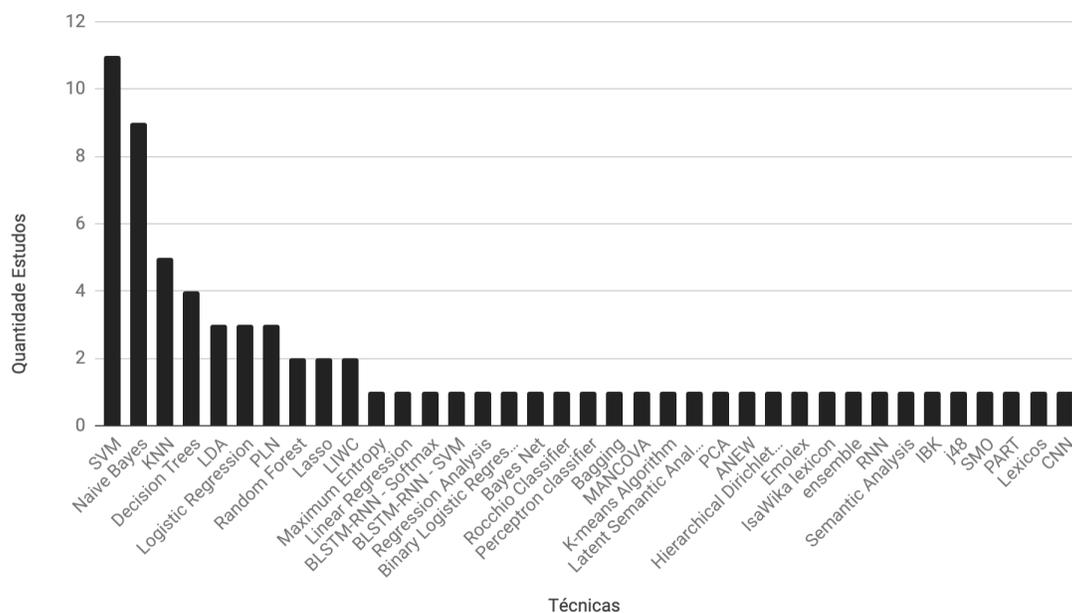


Figure 4: Técnicas identificadas nos estudos primários

Em [11] foram coletados comentários do *Reddit* que continham termos como "feel useless", "hang myself", métodos de suicídio, nomes de medicamentos, entre outros. Após isso os pesquisadores

inspecionaram manualmente os dados para garantir a relevância dos mesmos.

Aldarwish e Ahmad [12] classificaram 2073 postagens como deprimidas e 4700 como não deprimidas, após isso as postagens

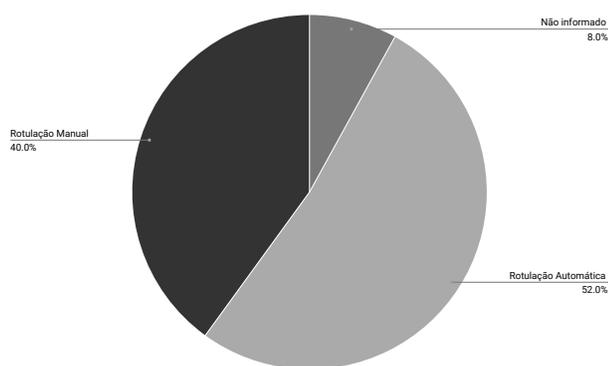


Figure 5: Tipos de rotulação identificados nos estudos primários

deprimidas foram categorizadas entre os nove sintomas definidos pelo (DSM-V).

Para criar a base de dados Tao et. al [22] realizaram buscas manuais de termos referentes a depressão no *Twitter* e em vocabulários *online*. Após a etapa de coletar dados, a base gerado foi analisada por vários psicólogos. Leiva e Freire [23] utilizaram a base de dados fornecida por [31], onde estes realizaram a coleta de postagens de diferentes usuários do *Reddit*. Após a coleta, os dados foram rotulados de forma manual onde os possíveis rótulos eram: se um usuário estava ou não com risco de possuir depressão.

O estudo de Briand, Almeida e Meurs [25] utilizou a base de dados disponibilizada por [32], a qual foi criada recuperando informações de usuários do *Reddit*, onde parte do conteúdo da base foi rotulado manualmente por psicólogos, entre usuário com risco (deprimido) ou sem risco (não deprimido).

No estudo de Tung e Lu [26] foi coletado um total de 18.000 postagens em chinês do fórum PTT. Dessas, 724 foram selecionados para treinamento e teste, essas postagens foram segmentadas e marcadas. Em seguida, foram rotuladas por três alunos de mestrado do *Institute of Behavioral Medicine College of Medicine, National Cheng Kung University*, como verdadeiras ou falsas referente a possuírem tendência à depressão.

A base de dados do estudo de Boukil et. al [27] foi criado a partir de postagens escritas em árabe, coletadas de diversos sites (*Facebook*, *Twitter*, *blogs* e fóruns). Todos os dados coletados foram manualmente analisados e rotulados nas seguintes classes: notas com indicação suicida e notas sem indicação suicida.

Kang, Yoon e Kim [29] coletaram 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*". Após a coleta os dados foram rotulados por três pessoas com experiência no uso de redes sociais. Para cada postagem era possível aplicar os seguintes rótulos: negativo, neutro ou positivo. Foram incorporados a base de dados somente postagens com os rótulos designados por duas pessoas no mínimo, correspondendo a 89,9%.

- **Rotulação Automática:**

Deshpande e Rao [8] separaram 80% do *dataset* para treinamento, e 20% para a realização dos testes. Os *tweets* coletados para a base de treinamento deveriam conter palavras como "*depressed*", "*suicide*" e "*hopeless*". Já para a base de testes os *tweets* foram coletados de forma aleatória.

No estudo de Nguyen et. al [5] foi criada uma base clínica, utilizando a "busca de comunidades por interesse" disponível no *LiveJournal*, utilizando o termo "*depression*" como entrada da busca. O resultado foram 24 comunidades, que somadas, continham 38,401 postagens. Também criou uma base de controle utilizando a mesma busca supracitada, utilizando como entrada as cinco categorias mais populares do blog, o resultado desta foi busca foi de 229,563 postagens contidas em 23 comunidades.

O estudo de Rosa et. al [13] filtrou sentenças providas de postagens do *Facebook* que continham expressões como "odeio minha vida", "me sentindo triste", "estou estressado", entre outras. Também utilizou um conjunto de sentenças positivas para classificar como "não depressão". A partir disso, 27,308 mensagens foram coletadas, onde 23,7% é referente à depressão, 26,18% é referente à estresse, por fim 50.1% são mensagens que não tem relação com estresse e nem com depressão.

Para identificar os potenciais *tweets* emocionais, Larsen et. al [15] utilizou os vocabulários léxicos ANEW e o LIWC, onde o primeiro fornece um conjunto de avaliações emocionais para uma grande quantidade de palavras em inglês, e o segundo é uma ferramenta proprietária que auxilia na extração de características das palavras, e na identificação da polaridade que as mesmas apresentam, assim avaliando o grau que cada texto tem em diferentes categorias.

Dao et. al [16] utilizou os conjuntos de humor definidos nas postagens do *LiveJournal* para catalogar a base, e a dividiu em três grupos: depressão com 38,400 postagens, autismo que contém 10,000 postagens e geral que possui 220,000 postagens.

No estudo de Hussain et. al [17] foi realizada uma busca nos últimos 100 *tweets* de 2,151 usuários, e as últimas 100 postagens de 515 usuários públicos do *Facebook*, após isso a base foi rotulada utilizando vocabulários léxicos de duas línguas diferentes Emolex(inglês) e IsaWika Lexicon(Tagalo). Esse vocabulários léxicos foram utilizados para determinar a tristeza de uma pessoa.

Nikhita e Srinivasan [18] identificaram usuários do *Twitter* com depressão, com auxílio médico. Após essa identificação a base de treinamento foi montada com os *tweets* do período de julho de 2016 a janeiro de 2017 dos usuários supracitados, bem como dos seus seguidores. O estudo de Islam et. al [19] coletou os dados do *Facebook* com o auxílio do NCapture. Após a coleta, foi utilizado o vocabulário léxico LIWC para calcular o valor através do padrão escalas do LIWC2015.

A base de dados do estudo de Wu et. al [20] é constituída com a co-relação de postagens coletadas do *Facebook* de universitários de Tawian que responderam o questionário *online* CES-D. Os resultados de triagem dos testes do CES-D. E ainda com dados de ambiente disponíveis ao público tais como: histórico de acidentes de trânsito, acidentes em geral, densidade da população, entre outros. Na base de dados deste estudo cada usuário é rotulado como deprimido ou não de acordo com a pontuação total atingida no CES-D, onde atingir mais de 60 pontos significa uma maior frequência de experiências depressivas.

Hasan, Rundensteiner e Agu [21] utilizaram palavras simples para realizar a busca de informações chamadas também de uni-gramas. Após a coleta das informações, para selecionar e manter no conjunto de dados apenas palavras com sinais de emoções, foram utilizados diferentes dicionários léxicos tais como: LIWC, ANEW, AFINN. Este último foi criado para incluir uma nova lista de palavras, focado especialmente para microblogs.

O estudo [6] utilizou a "busca de comunidades por interesse" presente no *LiveJournal*. Foram selecionadas apenas as comunidades que possuíam no mínimo 200 postagens e pelo menos uma atualização no mês anterior a busca. Ao final foram selecionadas 24 comunidades com um total de 38.401 postagens, as quais foram agrupadas nos seguintes subgrupos: depressão, transtorno bipolar, auto-mutilação, luto e suicídio. Para que criar a base de dados de forma balanceada foram escolhidos aleatoriamente 1.000 postagens de cada grupo.

Os autores Birjali, Beni-Hssane e Erritali [24] optaram por buscar postagens do *Twitter* e as rotular de forma automática utilizando para isso o método de similaridade semântica *WordNet* tendo como base um "vocabulário da depressão" definido previamente.

Para criar a base de dados do estudo de Ma, Wang e Zhang [28] foram monitorados postagens do *Twitter* contendo a palavra "depressão". Postagens providas de contas de profissionais de saúde mental. E ainda *blogs* de depressão. Com isso totalizando a base de dados com mais de 54 milhões de postagens referentes a depressão.

Observa-se ainda que dentre todas as bases que foram criadas pelos estudos selecionados, apenas quatro foram disponibilizadas, o estudo de Hassan et. al [7] utilizou o *dataset 20newsgroups* que é popular para experimentos de técnicas de aprendizado de máquina, como classificação e agrupamento de textos. O estudo de Larsen et. al [15] disponibiliza sua base de dados por meio da API¹. A base de dados utilizada pelos estudos [23] e [25] está disponível para fins de pesquisa sob contratos de adequação de usuário².

3.3 Análise e Discussão

Durante a execução deste MS foram encontrados dois estudos com o mesmo propósito. O primeiro é um *Survey* [33] publicado no *Springer* no ano de 2019. O segundo, consiste em uma Revisão Sistemática [34] publicada em 2018 no *Scopus*.

Em [33] fez uma revisão de estudos relacionados a técnicas e algoritmos de *Data Mining* em saúde mental até Março de 2018. O enfoque deste estudo era mais amplo, pois abrangia além da depressão, doenças como demência, Alzheimer, Esquizofrenia, e ainda saúde mental em geral, além de não filtrar somente estudos que se propuseram a análise textual das doenças, onde condicionou na inclusão de estudos que realizaram a análise de exames cerebrais de imagens. Entre os resultados referentes a depressão, os autores identificaram que SVM e NB foram os algoritmos mais utilizados para compor as abordagens dos estudos encontrados.

Alonso et al. [34] fez a coleta de estudos também de forma ampla, buscando além de estudos que tinham como ferramenta de estudo dados textuais. Estudos que buscavam verificar se alterações cognitivas influenciavam na forma de utilizar a linguagem. Os autores criaram uma seção para descrever as ferramentas utilizadas pelos

estudos, onde foram identificados dicionários léxicos como LIWC e CELEX, algoritmos de redução de dimensionalidade como LDA. *Frameworks* como *Natural Language Toolkit* (NLTK) e *Scikit-Learn*. Questionários como CES-D e *The Beck Depression Inventory* (BDI). E ainda abordagens de AM supervisionadas como SVM, *Hidden Markov Models* (HMM), *Recurrent Neural Networks* (RNN) e *Long short-term memory Neural Network* (LSTM).

Visando uma melhor contextualização das resposta de cada QP, a seguir são apresentados os objetivos de alguns dos estudos primários selecionados. Essa análise visa compreender quais técnicas têm sido empregadas, as bases de dados que são utilizadas e como os estudos alcançaram seus respectivos resultados na utilização de AS. Por fim, é importante ressaltar que algumas dessas informações não estavam disponíveis nos estudos, assim o que será apresentado será de acordo com a interpretação dos condutores do MS.

Diferentes estudos [7], [8], [12], [15], [17], [18], [21], [22], [24] e [28] utilizaram o *Twitter* como fonte de dados para criação das bases de dados. Dentre tais estudos, é importante destacar o estudo de Nikhita e Srinivasan [18], o qual teve como objetivo identificar se existe uma diferenciação no comportamento *online* entre usuários potencialmente deprimidos, com relação a usuários normais. Os autores observaram que existem desvios significativos no comportamento dos usuários deprimidos, alguns exemplos são: atividade *online* reduzida e noturna, uso de emoções negativas nos textos e uso de pronomes auto-focados. Para detectar os sinais sintomáticos da depressão, foram examinados sinais como: estilos linguísticos, onde pode ser destacado aumento do uso de pronomes pessoais associados e derivados, e redução de pronomes de terceira pessoa ou que possuem conotação coletiva; sinais emocionais; interação do usuário e localização geográfica.

Os estudos [9], [12], [13], [14], [17], [17] e [20] utilizaram postagens do *Facebook* para compor suas respectivas base de dados de treinamento e testes. Dentre eles destaca-se o estudo de Rosa et. al [13] por ser o único estudo dentre todos os 25 que utilizou uma fonte de dados em português. Os autores realizaram o desenvolvimento de uma aplicação *web* que tem como objetivo servir como um sistema de recomendação para aliviar sintomas da depressão, para isso foi criado uma base de textos a partir de ontologias. O sistema inclui uma fase de monitoramento que visa detectar depressão e estresse nos usuários. A operação proposta é de coletar sentenças e filtrar por meio de AM para detectar sinais de depressão e estresse. As sentenças filtradas então são classificadas de acordo com a sua intensidade, podendo ser: extremo, intermediário e inferior. O sistema de recomendação se baseia nessa intensidade para definir quais itens irá sugerir para o usuário.

Somente quatro estudos [12], [5], [16] e [6] utilizaram o *LiveJournal*. O estudo que merece destaque é o de Nguyen et. al [5], que visa demonstrar o potencial que as redes sociais possuem para identificar e monitorar a depressão em usuários. Para isso, foram investigadas comunidades de depressão online, visando encontrar os fatores diferenciais com relação a outras comunidades online. Foi identificado que aspectos como afeto e processos psicolinguísticos afetam o conteúdo escrito e o estilo da escrita em usuários depressivos.

O estudo de Leiva e Freire[23] é o que mais se destaca dentre os estudos ([11], [23] e [25]) que utilizaram o *Reddit* como fonte

¹<http://wefeeel.csiro.au/#/api>

²<http://tec.citius.usc.es/ir/code/dc.html>

de dados. Neste estudo o objetivo é melhorar o estado da arte referente a detecção antecipada de depressão em redes sociais. Os experimentos foram executados em um conjunto de dados criado com postagens do *Reddit*. Os textos foram extraídos de maneira mais precisa se comparado com as abordagens que utilizam apenas Term Frequency–Inverse Document Frequency (TF-IDF). Foram explorados o comportamento de técnicas clássicas e modernas de AM para prever de uma melhor forma os casos positivos. Ainda foi analisado a influencia de algoritmos genéticos nos resultados.

O estudo de [10] utilizou o Sina micro blog, já que o objetivo deste estudo era analisar dados textuais escritos em chinês, e esta fonte de dados é uma especie de twitter existente na china , onde a quantidade majoritária das informações postadas são em chinês.

Notas em árabe foram utilizados por Boukil et. al [27] para compor sua base de dados, a justificativa da escolha desta fonte de dados foi a dificuldade de encontrar informações referentes a depressão e ao suicídio escritas em árabe, já que em países árabes o tabu referente a estes problemas é maior que no restante do mundo. O objetivo dos autores é criar um sistema automático para extrair e prever a intenção de suicídio de um indivíduo, com base em mensagens e notas textuais do mesmo. Para isso, foi utilizado o método TF-IDF para reformular as notas textuais na forma que o sistema possa reconhecer e entender. Posteriormente foi criado um modelo utilizando Convolutional Neural Network (CNN) para reconhecer e classificar notas que apresentam potencial suicida e notas que não apresentam este potencial.

Em Hassan et. al [7] utilizaram a base de dados *20newsgroups* para testar a assertividade de seus resultados, já que os mesmos tinham como principal objetivo realizar a comparação entre as técnicas: SVM, NB e *Maximum Entropy* (ME) para detectar depressão em diferentes bases de dados. Para combinar os múltiplos classificadores, este estudo utilizou o modelo de voto majoritário, que consiste em escolher a classe que obteve o maior número de classificações entre as técnicas utilizadas no processamento dos dados.

O fórum de discussões PTT foi utilizado por Tung e Lu em [26], onde os autores utilizam técnicas de PLN, recuperação de informações e mineração de textos, com o objetivo de transferir conhecimento especializado e, conseqüentemente, aprimorar a análise e dicção da tendência de depressão em postagens na *internet*. Este estudo ainda utilizada léxicos para realizar a análise, e uma de suas contribuições é a expansão de um léxico de eventos negativos de forma automática.

Em [28] foram utilizados *blogs* de depressão em geral, onde o objetivo dos autores era de extrair e resumir fatores incomuns, mas que sejam relevantes para a identificação de sintomas depressivos, com isso aumentando a possibilidade da ferramenta criada pelos mesmos de conseguir identificar de forma correta perfis considerados depressivos.

Vocabulários online foram utilizados por Tao et. al [22] para aumentar a quantidade de termos que estão relacionados a depressão. Onde estes termos auxiliaram na análise das informações publicadas pelos usuários do *Twitter*. Foram utilizadas técnicas de PLN, mais especificamente análise semântica. A abordagem proposta pelos autores possui ainda o objetivo de oferecer a assistentes sociais a capacidade de acessar potenciais pessoas depressivas que precisam de ajuda no estágio inicial da depressão.

O estudo de Wu et. al [20] foi o único que utilizou dados do CES-D para auxiliar na identificação dos perfis depressivos dos participantes do estudo, onde foram relacionadas as respostas dos fornecidas pelos mesmos no questionário com as informações coletadas das redes sociais.

Os estudos [7], [12], [17], [22] e [28] utilizaram mais de uma fonte de dados para criar suas respectivas bases de dados. O motivo disso pode estar ligado ao fato de que apenas uma fonte de dados pode ser insuficiente para conseguir generalizar um perfil depressivo. Ou ainda com o objetivo de validar a abordagem criada em uma fonte de dados diferente da de treinamento, como no caso de Hassan et. al [7], no qual o *20newsgroups* foi utilizado para testar o modelo.

Fazer o uso de tecnologia para ajudar pessoas que possuem depressão é de grande relevância, pois nos últimos anos a quantidade de portadores desta doença aumentou em números alarmantes [1]. Os estudos encontrados durante a execução deste MS provaram ser possível identificar sintomas depressivos em dados textuais, portanto é possível alertar indivíduos que não tem o conhecimento de possuir a patologia e conseqüentemente não procuram auxilio profissional. Após a análise dos resultados, observou-se que as duas técnicas mais recorrentes entre os estudos foram NB e SVM, os quais foram utilizadas por 12 estudos ([7], [8], [12], [13], [14], [19], [21], [22], [23], [25], [25] e [29]). A maioria dos estudos relataram resultados satisfatórios utilizando as técnicas citadas na Tabela 1, pois em geral as mesmas atingiram uma boa acurácia durante as classificações realizadas.

Zuorba et al. [17] destaca-se como o estudo que obteve o melhor resultado entre os demais. Este estudo utilizou um modelo de Regressão Logística (*Logistic Regression*) que foi treinado com os vetores de características gerados a partir dos léxicos. O *Latent Dirichlet Allocation* (LDA) foi utilizado para melhorar a capacidade de discriminação entre as classes, reduzindo a dimensionalidade por meio da geração de tópicos.

Vale ressaltar que [14] é uma análise comparativa das técnicas mencionadas, e por isso possui um elevado número de técnicas relacionadas. Com a análise dos resultados, foi possível observar que [11], [17] e [29] (LDA) e o [15] (PCA) utilizaram técnicas de redução de dimensionalidade das bases de dados. Essas técnicas são utilizadas para melhorar a capacidade de discriminação entre as classes existentes na base de dados.

3.4 Ameaças à validade

Principais ameaças à condução de um MS está relacionada ao fator humano que pode introduzir viés no processo de seleção dos estudos primários, além de provocar imprecisão na extração dos dados. Neste MS foram identificadas algumas ameaças à validade, bem como possíveis melhorias para minimizar ou extinguir essas ameaças.

A primeira ameaça a ser discutida é o número de pesquisadores. Somente um pesquisador realizou este MS, aumentando o risco de vieses no trabalho. Com o objetivo de minimizar este risco, os resultados encontrados, bem como todas as fases do MS foram acompanhadas e validadas pelo orientador deste trabalho.

A segunda ameaça está relacionada com a criação e adaptação da *string* de busca. Como as palavras e as expressões que compõem são derivadas a partir das QPs, a correta construção desta é vital

para a efetividade da pesquisa. Da mesma forma, a adaptação da *string* para cada base pode mudar o foco da pesquisa caso não seja feita de forma precisa. Para mitigar esta ameaça, foi solicitado a um especialista que avaliasse a *string* para validá-la e melhorar sua efetividade. Por fim, a última ameaça é referente à extração de dos dados. Neste caso, devido a omissão de informações importantes em vários estudos, houve a necessidade do pesquisador buscar outras publicações relacionadas a fim de auxiliar na análise dessas informações

4 CONSIDERAÇÕES FINAIS

Utilizar AS para identificar traços depressivos em usuários de redes sociais mostra-se uma atividade promissora, pois os estudos selecionados neste MS demonstraram que pessoas que possuem depressão, tristeza excessiva, ou sinais depressivos, tendem a apresentar um comportamento *online* diferente se comparado com o comportamento de usuários que não possuem tais características. Alguns aspectos desse comportamento são: (i) forma de se manifestar; (ii) estilos linguísticos tais como, aumento da utilização de pronomes auto-focados e expressões negativas; (iii) mudança no horário de utilização das redes sociais, onde o mesmo torna-se em grande maioria noturno, entre outros.

A principal contribuição desse trabalho está na obtenção de uma visão geral das possíveis técnicas e tipos de rotulação da base de dados utilizadas para análise de sentimentos e identificação de perfis depressivos em redes sociais. Por fim, os resultados alcançados após a execução deste MS, demonstram ainda, que apesar da depressão ser um tema atualmente muito discutido no âmbito global, e o uso de AS vir se provando ao longo dos anos como uma excelente maneira de analisar e entender opiniões dos usuários das redes sociais, poucos estudos tem sido realizados envolvendo essas duas áreas de pesquisa nos últimos anos, principalmente quando se trata da análise dos dados em português.

Como trabalhos futuros, pretende-se complementar os resultados obtidos nesse MS por meio da busca de novos artigos em outras bases eletrônicas, por meio da busca manual em anais de congressos na área e por meio de referências.

REFERENCES

- [1] World Health Organization et al. Depression and other common mental disorders: global health estimates. Technical report, World Health Organization, 2017.
- [2] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. Systematic mapping studies in software engineering. In *12th International Conference on Evaluation and Assessment in Software Engineering*, pages 68–77, 2008.
- [3] B. Kitchenham, R. Pretorius, D. Budgen, P. Brereton, M. Turner, M. Niazi, and S. Linkman. Systematic literature reviews in software engineering - a tertiary study. *Inf. Softw. Technol.*, pages 792–805, 2010.
- [4] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil. Lessons from applying the systematic literature review process within the software engineering domain. *Systems and Software*, 80:571–583, 2007.
- [5] T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk. Affective and content analysis of online depression communities. *IEEE Transactions on Affective Computing*, 5(3):217–226, 2014.
- [6] Thin Nguyen, Bridianne O’Dea, Mark Larsen, Dinh Phung, Svetha Venkatesh, and Helen Christensen. Using linguistic and topic analysis to classify sub-groups of online depression communities. *Multimedia Tools and Applications*, 76(8): 10653–10676, 2017.
- [7] A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee. 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, 2017.
- [8] M. Deshpande and V. Rao. Depression detection using emotion artificial intelligence. In *International Conference on Intelligent Sustainable Systems*, pages 858–862, 2017.
- [9] M. R. Islam, A. R. M. Kamal, N. Sultana, R. Islam, M. A. Moni, and A. ulhaq. 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, 2018.
- [10] X. Wang, C. Zhang, and L. Sun. An improved model for depression detection in micro-blog social network. In *2013 IEEE 13th International Conference on Data Mining Workshops*, pages 80–87, 2013.
- [11] J. H. K. Seah and K. Jin Shim. Data mining approach to the detection of suicide in social media: A case study of singapore. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 5442–5444, 2018.
- [12] M. M. Aldarwish and H. F. Ahmad. Predicting depression levels using social media posts. In *2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS)*, pages 277–280, 2017.
- [13] R. L. Rosa, G. M. Schwartz, W. V. Ruggiero, and D. Z. Rodriguez. A knowledge-based recommendation system that includes sentiment analysis and deep learning. *IEEE Transactions on Industrial Informatics*, 15(4):2124–2135, 2019.
- [14] A. Noureen, U. Qamar, and M. Ali. Semantic analysis of social media and associated psychotic behavior. In *International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, pages 1621–1630, 2017.
- [15] M. E. Larsen, T. W. Boonstra, P. J. Batterham, B. O’Dea, C. Paris, and H. Christensen. We feel: Mapping emotion on twitter. *IEEE Journal of Biomedical and Health Informatics*, 19(4):1246–1252, 2015.
- [16] B. Dao, T. Nguyen, S. Venkatesh, and D. Phung. Nonparametric discovery of online mental health-related communities. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–10, 2015.
- [17] Hussain D. Zoorba, Celine Louise O. Olan, and Anthonette D. Cantara. A framework for identifying excessive sadness in students through twitter and facebook in the philippines. In *ICBRA 2017*, 2017.
- [18] Nikhita Vedula and Srinivasan Parthasarathy. Emotional and linguistic cues of depression from social media. In *DH*, 2017.
- [19] Md. Rafiqul Islam, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M. Kamal, Hua Wang, and Anwaar Ulhaq. Depression detection from social network data using machine learning techniques. *Health Information Science and Systems*, 6(1): 8, 2018.
- [20] Min Yen Wu, Chih-Ya Shen, En Tzu Wang, and Arbee L. P. Chen. A deep architecture for depression detection using posting, behavior, and living environment data. *Journal of Intelligent Information Systems*, 2018.
- [21] Maryam Hasan, Elke Rundensteiner, and Emmanuel Agu. Automatic emotion detection in text streams by analyzing twitter data. *International Journal of Data Science and Analytics*, 7(1):35–51, 2019.
- [22] Xiaohui Tao, Xujuan Zhou, Ji Zhang, and Jianming Yong. Sentiment analysis for depression detection on social networks. pages 807–810, 2016.
- [23] V. Leiva and A. Freire. Towards suicide prevention: Early detection of depression on social media. In *Internet Science*, pages 428–436, 2017.
- [24] M. Birjali, A. Beni-Hssane, and M. Erritali. A method proposed for estimating depressed feeling tendencies of social media users utilizing their data. pages 413–420, 2017.
- [25] A. Briand, H. Almeida, and Marie-Jean Meurs. Analysis of social media posts for early detection of mental health conditions. pages 133–143, 2018.
- [26] C. Tung and W. Lu. Analyzing depression tendency of web posts using an event-driven depression tendency warning model. *Artif. Intell. Med.*, pages 53–62, 2016.
- [27] S. Boukil, F. El Adnani, L. Cherrat, A. E. El Moutaouakkil, and M. Ezziyyani. Deep learning algorithm for suicide sentiment prediction. pages 261–272, Cham, 2019. Springer International Publishing.
- [28] Long Ma, Zhibo Wang, and Yanqing Zhang. Extracting depression symptoms from social networks and web blogs via text mining. pages 325–330. Springer International Publishing, 2017.
- [29] Keumhee Kang, Chanhee Yoon, and Eun Yi Kim. Identifying depressive users in twitter using multimodal analysis. In *International Conference on Big Data and Smart Computing*, pages 231–238, 2016.
- [30] Xinyu Wang, Chunhong Zhang, Yang Ji, Li Sun, Leijia Wu, and Zhana Bao. A depression detection model based on sentiment analysis in micro-blog social network. In *PAKDD Workshops*, 2013.
- [31] D. E. Losada and F. Crestani. A test collection for research on depression and language use. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, 2016.
- [32] D. E. Losada, F. Crestani, and J. Parapar. erisk 2017: Clef lab on early risk prediction on the internet: Experimental foundations. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 346–360, 2017.
- [33] Dirk Johannßen and Chris Biemann. Between the lines: Machine learning for prediction of psychological traits - a survey. In *Machine Learning and Knowledge Extraction*, pages 192–211, Cham, 2018.
- [34] S. G. Alonso, I. de la Torre-Díez, S. Hamrioui, M. López-Coronado, D. C. Barreno, L. Nozalea, and M. Franco. Data mining algorithms and techniques in mental health: A systematic review. *Journal of Medical Systems*, 42(9):161, 2018.