Titla-Tlatelpa et al. *EPJ Data Science* (2021) 10:54
https://doi.org/10.1140/epjds/s13688-021-00309-3

EPJ Data Science
a SpringerOpen Journal

EPJ.org

REGULAR ARTICLE

Open Access

A profile-based sentiment-aware approach for depression detection in social media

José de Jesús Titla-Tlatelpa[1], Rosa María Ortega-Mendoza[2*], Manuel Montes-y-Gómez[1] and Luis Villaseñor-Pineda[1]

*Correspondence:
rosa.ortega@upt.edu.mx
[2]Universidad Politécnica de Tulancingo, Calle Ingenierías #100, Huapalcalco, Tulancingo de Bravo, Hidalgo, México
Full list of author information is available at the end of the article

Abstract

Depression is a severe mental health problem. Due to its relevance, the development of computational tools for its detection has attracted increasing attention in recent years. In this context, several research works have addressed the problem using word-based approaches (e.g., a bag of words). This type of representation has shown to be useful, indicating that words act as linguistic markers of depression. However, we believe that in addition to words, their contexts contain implicitly valuable information that could be inferred and exploited to enhance the detection of signs of depression. Specifically, we explore the use of user's characteristics and the expressed sentiments in the messages as context insights. The main idea is that the words' discriminative value depends on the characteristics of the person who is writing and on the polarity of the messages where they occur. Hence, this paper introduces a new approach based on specializing the framework of classification to profiles of users (e.g., males or women) and considering the sentiments expressed in the messages through a new text representation that captures their polarity (e.g., positive or negative). The proposed approach was evaluated on benchmark datasets from social media; the results achieved are encouraging, since they outperform those of state-of-the-art corresponding to computationally more expensive methods.

Keywords: Depression detection; Author profiling; Sentiment analysis

1 Introduction

Depression is a common mental health problem that severely impacts our society. More than 264 million people of all ages suffer from depression worldwide[1] affecting seriously their quality of life and physical health. Unfortunately, in severe cases, depression can even result in suicide [1]. The relevance of this health problem has motivated the development of computational tools for the automatic detection and monitoring of people suffering from this mental disorder [2]. Several studies have shown that people who suffer from this disorder alter their written language and communicate differently in both style and content [3]; for example, high self-focused attention and high usage of negative words can reveal signs of depression [4]. The link between language and the psychological state of

[1]https://www.who.int/news-room/fact-sheets/detail/depression

Springer

people has led to the exploration of data from social networks for the automatic detection of depression, aiming to take advantage of the large amount of information generated by people through these media, in which they tend to express their thematic interests, experiences, desires, concerns, etc. Accordingly, recent evaluation forums have focused on detecting people suffering from depression through the analysis of their social media posts [5–7].

Traditionally, the automatic detection of depression has been tackled as a supervised text classification problem. Recent works have addressed it using neural networks fed by word embeddings [8, 9], but also have considered a diversity of text representations such as histograms of word categories (e.g., LIWC[2]), broad and fine-grained emotions [10, 11], latent topics extracted with LDA [12, 13], and traditional *word-based representations* [14]. The latter have shown to be useful and competitive for this task, indicating that words act as linguistic markers of depression. Although their good results, we argue these representations can be enriched by additional *context information*, which, although usually not indicated explicitly, can be inferred from the posts. In particular, the proposed approach exploits demographic characteristics of the users (inferred from the texts) and the polarity of the messages on the basis of two key ideas: (*i*) people belonging to the same group (e.g., males or females) tend to manifest and express their depression similarly but differently to other groups and, (*ii*) words are important markers of depression, but their correct interpretation depends on the polarity, positive or negative, of their usage context.

Some previous works have pointed out differences related to the prevalence of mental disorders regarding users' traits, such as gender [15, 16], age [15], and occupation [17]. For example, depression is more common in women, especially around puberty [18]. Furthermore, it is expected that different groups of people manifest their depression differently, because each group has its own interests and way of communicating; take for instance the case of women and men, or young and mature persons. Supported on this idea, we propose an approach that *specializes classifiers according to the users traits*. On the other hand, recent studies indicate that people's psychological well-being is associated with the type of relationship perceived between positive and negative affect, as independent or bipolar opposites [19, 20]. Inspired by these ideas, we consider that to distinguish depression, it is essential to analyze the contexts of use of words, both positive and negative, in a dual way. For example, depressed and non-depressed users can talk about the same topic (e.g., about their partners), but the polarity of their contexts can be very different, negative (e.g., a separation) or positive (e.g., a nice experience), respectively. Hence, we propose using *a dual text representation* which allows encapsulating both polarities to provide the classifiers with the possibility of relating them and finding discriminative patterns.

In general, the proposed approach jointly uses the words, the polarities of their posts, and the profile traits of the users to distinguish those suffering from depression. In other words, it is aiming to discover word-sentiment patterns associated with different groups of people suffering from depression, such as the negative usage of the word "calories" in the case of depressed women, or the high occurrence of sexuality issues, for example, the word "virgin", in negative posts for the case of young depressed users. Summarizing, the main contributions of this work are threefold: (*i*) a new approach to the detection of depression in social media, which applies specialized classifiers for different groups of users

[2] https://liwc.wpengine.com/

and takes into account the polarities of their posts; (*ii*) a new dual bag-of-words representation, which captures and distinguishes the occurrences of the words in both positive and negative contexts; and (*iii*) an in-deep analysis of the role of profiling and polarity information in the depression detection task.

The rest of this paper is organized as follows. Section 2 provides a review of the related work on depression detection on social media. Section 3 introduces the proposed approach. Section 4 explains the proposed text representation. Section 5 presents the experimental settings. Section 6 shows the experiments and results. Section 7 exposes an analysis and discussion of the obtained results. Finally, our conclusions and future work directions are drawn in the last section.

2 Related work

2.1 Depression detection on social media

Social networks are increasingly used to share daily activities; moreover, they are also used to connect with others as a form of social support on health issues [21]. Under this scenario, computational approaches have leveraged the information from social media to study the depression as well as to detect users suffering from depression. For example, some studies use representations based on BoW [22–24] to make the detection at user and post levels. These kinds of representations allow to easily measure and compare the utility of word n-grams to identify depression in posts. Representations based on topics have also been explored. For example, [25] analyzed online health forums to identify changes in the language and topics to which depressive users are mostly associated. Some other works based on topics have explored the use of resources and techniques such as LIWC and LDA [26–28]. In this regard, the combination of manually and automatically generated topics has shown good performance [13]. Recently, due to the relevance of the problem, some evaluation forums such as eRisk[3] have motivated the development of computational approaches to face the early detection of social media users suffering from depression [5, 6]. Different architectures have been evaluated through the editions of this forum, from new methods based on word representations [29] to complex architectures using deep learning architectures [30], reaching results around F1 = 0.65 over the positive class, which suggest that distinguishing depressed from non-depressed social media users remains as a major challenge.

As can be observed, various approaches and methods have been proposed to predict the depression disorder from a computational perspective on social media. The idea behind all of them is to help in clinical care. In this regard, recently, Chancellor and De Choudhury [31] studied some issues of construct validity that could inhibit reproducibility and extension into practical and clinical domains, for example, issues of reporting practices. Accordingly, the authors provided some interesting recommendations to addresses these challenges.

2.2 The role of profiling traits for depression detection

From a psychological perspective, the role of demographic factors in mental illness has been studied [32, 33]. Particularly, several works have analyzed the relationship between patients' profile attributes (e.g., age, gender, and personality traits) and the manifestation

[3] https://erisk.irlab.org/

of their depression [34]. Most of them have found clear differences among distinct groups of people, particularly between men and women [35–37]. For example, differences in depression by gender have been explored by considering social roles, norms of culture, family environment, and biological factors [3]. Some works have also explored differences in the relationship between social media use and depressive symptoms in the child and adolescent population, finding a significant correlation between usage patterns and depressive symptoms in young people [38].

From a computational perspective, some works approaching the automatic detection of depression in social media have used the age and gender of users as classification features. For example, in [33] the detection of users suffering from depression was carried out considering *only* profile attributes. The results reported using age and gender attributes were 15% better than random guessing, suggesting that they provide relevant information for the task at hand. In [22] these two demographic features were used in conjunction with word n-grams, part of speech (POS) tags, emoticons, sentiment polarities, and LIWC categories. Although the results obtained were not conclusive regarding their relevance to the task, because they were used in combination with many other features, the authors suggested they contribute for the good performance of their model. Similar to these two approaches, we also take advantage of the profile information of users, but not as extra features, instead we aim to infer profiles for specializing the classification process according to them.

2.3 Emotions in depression detection

Emotional information has been mainly studied by means of discrete emotions (e.g., sadness and anger), but also by its polarity in a dichotomous scale of positive and negative values, characterizing statements into positive or negative expressions mutually exclusive [39, 40]. From a psychological perspective, emotions are aspects which help to diagnose depression [41]; in particular negative affect has been associated with the depression disorder [42, 43]. Recently, the severity of depressive symptoms was associated with a more inverse relationship between positive and negative affect (i.e., high bipolarity), mainly because individuals with depressive symptoms present difficulty on regulating emotions, leading to a reduction of emotional complexity [19, 20]. In this regard, but in a slightly different direction, [44] studied the relationship between distorted thinking and depression, concluding that individuals with depression tend to exhibit higher levels of cognitive distortions, such as the dichotomous, catastrophic and disqualifying reasoning of the positive.

From a computational perspective, some studies have examined the use of representations based on emotions. For example, [10] considered the frequency of occurrence of the main emotions in the users' posts to identify those suffering from depression, showing their advantage over the exclusive use of linguistic characteristics. Then, [11] went a step further by using a bag of sub-emotions (BoSE) representation, in which posts are represented by a frequency histogram of fine-grained instead of broad emotions, allowing to capture more specific topics expressed by depressed users. These two recent contributions show that the emotional tone of the information is relevant for the detection of depression. Following this idea, but different from these works, here we explore in a dual way the positive and negative valences of the sentiments for establishing the discriminative value of words for revealing depression traits.

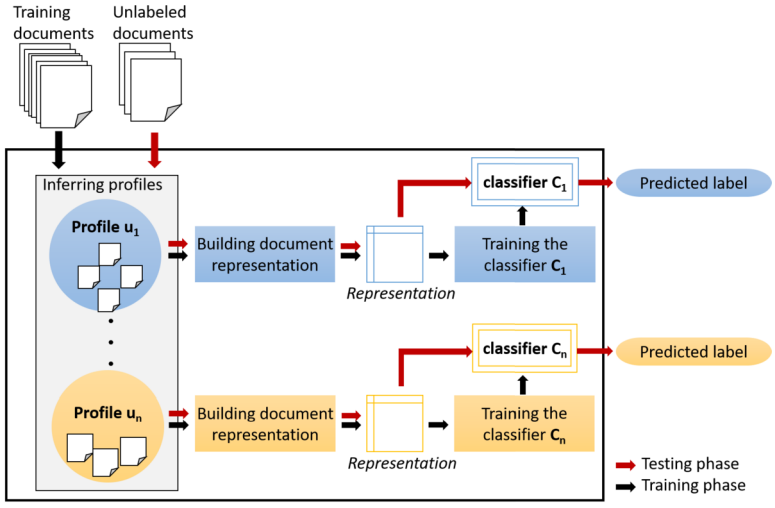


Figure 1 General approach for depression detection using specialized classifiers for the different users' profiles

3 Specializing classifiers according to users' profiles

As previously stated, the task of depression detection in social media has been addressed as a supervised, binary, text classification problem, which goal is to learn a classifier that categorizes the users, described by their post histories, into one of two possible categories, a user suffering from depression or a non-depressed user.

Inspired by the idea that people belonging to the same group manifest and express their depression similarly, we propose to build specialized classifiers for the different groups of users, defining these groups according to some of their traits. For example, to build independent classifiers for male and female users, or for young and senior users. The proposed approach is depicted in Fig. 1. In it, each user is described by a single *document*,[4] which contains all her/his posts, and each classifier c_i is specially trained to predict depression exclusively on the group of users corresponding to profile u_i. Accordingly, each unlabeled document (i.e., new user) will be evaluated only by the classifier specialized in its respective profile.

It is worth mentioning that *the proposed approach is general*, and it does not depend on the particular traits used to separate the users, nor on how these are determined, which could be manually or automatically. It can also be used in combination with any document representation. For the experiments, we consider gender and age traits automatically inferred from the texts, and a novel dual word-sentiment representation. For details refer to Sects. 5.5 and 4, respectively.

4 Bag of polarities: a dual word-sentiment representation for depression detection

Term-based representations, such as the bag of words (BoW), are commonly used in text classification tasks, showing satisfactory results in most of them. However, for complex tasks, which require finer discrimination between classes, their main drawback is that

[4]Each user has a post history, stored in a single document. Thus, from here on we will use "document" to refer to the set of posts of a given user.

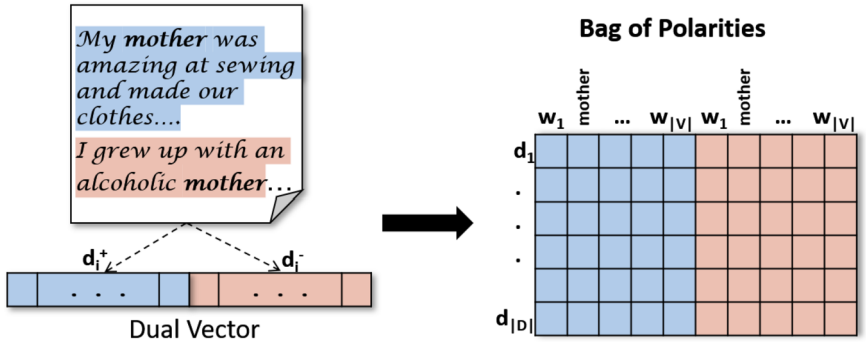


Figure 2 The bag of polarities representation; in it each document from the collection is represented by a dual vector that captures the occurrence of words in both positive and negative contexts

they do not capture information of the contexts of the words. To address this limitation, different types of contextualized word embeddings (e.g., ELMo, Flair, and BERT) have been recently proposed [45, 46]. Their idea is to add syntactic and semantic information to the words' representations aiming to dynamically capture their meaning. They have achieved outstanding results in several text processing tasks, but on the contrary, they have reduced the interpretability of the results.

Being aware of the relevance of the interpretability of results and explicability of methods in mental health applications, we decided to continue working with term-based representations, but extending them with some kind of context information. As previously mentioned, we consider that the value of the words as linguistic markers of depression largely depends on the *polarity* of their contexts of use. For example, when users mention words related to their family or work, it is essential to know their contexts' polarity (positive or negative) before considering them as relevant signs for establishing a user as suffering from depression. Accordingly, we propose a BoW-based dual representation that allows capturing both positive and negative uses of all words, thus providing the classifiers with the possibility of exploring the relationship between both kinds of mentions. Figure 2 illustrates this representation, which we named as *Bag of Polarities* (BoP). As shown, its construction starts from the identification and separation of positive and negative posts, indicated in blue and red respectively. Then, considering the vocabulary of the posts of all users, a BoW-type representation is built, which maintains information about the occurrences in positive and negative contexts of each word, making the size of BoP double that of a traditional BoW. That is, each word is mapped to two different components of the representation space, see for example the case of the word "mother", which occurs in the blue as well as red sections of the representation, accounting for its positive and negative occurrences, respectively.

More formally, the BoP representation can be defined as follows:

Let $D = \{d_1, \ldots, d_{|D|}\}$ denote the set of documents (i.e., social media users in our case), and $V = \{w_1, \ldots, w_{|V|}\}$ its vocabulary; each document d_i is represented as vector a $\mathbf{d_i} = \mathbf{d_i^+} \| \mathbf{d_i^-}$, which results from the concatenation of its vectors of positive and negative contexts defined as:

$$\mathbf{d_i^+} = \langle v_{i,1}^+, \ldots, v_{i,|V|}^+ \rangle, \qquad \mathbf{d_i^-} = \langle v_{i,1}^-, \ldots, v_{i,|V|}^- \rangle, \tag{1}$$

Table 1 Description of datasets. Number of depressed (Dep) and non-depressed (NDep) users in the train and test partitions

Dataset	Training		Test		Total
	Dep	NDep	Dep	NDep	
Reddit	135	752	79	741	1707
Twitter	2626	5373	–	–	7999

where the $v_{i,j}$ values indicate the proportion of the occurrences of the word w_j in the positive and negative posts of document d_i. That is, if P_i, P_i^+, and P_i^- represent the set of posts in d_i and its respective subsets of positive and negative posts, where $P_i = P_i^+ \cup P_i^-$, then:

$$v_{i,j}^+ = \frac{f(w_j, P_i^+)}{f(w_j, P_i)}, \qquad v_{i,j}^- = \frac{f(w_j, P_i^-)}{f(w_j, P_i)}, \tag{2}$$

where $f(w, P)$ indicates the frequency of occurrence of a word w in the set of posts P.

As explained above, the BoP representation considers as extra information the polarity of the posts; similar to the approach presented in the previous section, it does not depend on how this information is determined, it could be inferred from the texts using any existing method for that purpose [47–49]. In our experiments, we used the procedure detailed in Sect. 5.4.

5 Experimental settings

5.1 Datasets

For evaluating the proposed approach, we used two benchmark datasets in English: a collection from Reddit users released in the context of the eRisk 2018 task [5] (hereafter denoted as *Reddit*), and a collection of Twitter users described in [50] (hereafter denoted as *Twitter*). Both collections were gathered by the respective authors using their own set of APIs. Table 1 summarizes their main statistics. The range of dates from the first submission of users to the last submission is around 500 and 30 days, for Reddit and Twitter, respectively. Both collections include users labeled as depressed and control (or non-depressed), and also their construction followed a similar approach: according to their authors, users were labeled as depressed if they explicitly mentioned (self-declaration) that were "diagnosed with depression", whereas non-depressed users correspond to users that never used the word "depressed" in their posts.

5.2 Text representations

All texts from the two collections were tokenized into unigrams, lower-cased, and the stop words and special characters were removed. To model the users' content, we considered the following two text representations:

- `Bag of Words (BoW)`: a standard BoW using unigrams and *tf-idf* weights. This representation acts as the baseline method in the experiments.
- `Bag of Polarities (BoP)`: the dual word-sentiment representation introduced in Sect. 4.

Both representations were built using the words from the training partition with the highest χ^2 values; 6000 and 10,000 words for Reddit and Twitter, respectively.[5]

[5]We also experimented with 500 to 10,000 words, with steps of 500, but our best results in the training partition were obtained with the referred numbers.

5.3 Classification and evaluation

During training we considered different classifiers, such as SVM, Random Forest, and a *Bagging of Decision Trees*; we decided using the latter as the base classifier in all the experiments because it showed the highest results. The hyperparameters of the model were automatically tuned using a 5-fold cross-validation over the training set and the sklearn's GridSearch algorithm. The search was focused on: (*i*) the number of trees (10, 20, 30, 40, 50) and (*ii*) their depth (3, 5, 6, 9, 10). Finding ideal values: 20 and 6 respectively. For evaluation purposes, in the case of Reddit, we ran the bagging algorithm five times and reported the average outcome on the test partition, whereas, in the case of Twitter, we applied a 5-fold cross-validation. In both cases, we report the F1 score over the positive class (i.e., depressed users), as it is the evaluation measure used in most previous works considering these two datasets.

5.4 Assignment of posts' polarities

The process to determine the posts' polarities is based on SentiWordNet [51], which is a resource that has three scores between −1 and 1 specified for each WordNet Synset, indicating how positive, negative, and objective are their words. In particular, for a given post, we estimate its polarity by averaging the scores corresponding to the synsets of each of its words. If the average is negative, then we assign the post a negative polarity, while if it is positive, we assign it a positive polarity. Table 2 shows the distributions of positive and negative posts in both collections, which in average correspond to 69% positive and 31% negative.

It is important to make three points: first, words missing in SentiWordNet were ignored; second, for words belonging to multiple synsets, we took the score of the synset corresponding to their most frequent meaning and that matches its part-of-speech label (to maintain the text connotation); third, posts having an average score equal to zero (i.e., neutral polarity) were not taken into account.

5.5 Inference of users' traits

As described in Sect. 3, we propose to specialize the classifiers according to different users' traits, particularly, gender and age, which allow us to differentiate between male and female users as well as between young and senior users. Given that the used collections do not contain information about the users' traits, we needed to apply a process to automatically infer them from their texts. We used two lexicons especially designed for this purpose [52], which include words weighted by their orientation; in the first lexicon positive and negative scores are associated with female and male respectively, whereas in the second lexicon positive scores indicate words commonly used by senior users, while negative weights are words more associated with young users.

The process to infer the users' traits is as follows: for a given document (i.e., a user), and using the two lexicons independently, we calculate the weighted sum of the scores of the words by their relative frequency in the document. The weighted sum directly indicates the user age; we labeled all users under the age of 25 as young, and the rest as senior.[6] On the other hand, the sign of the result indicates the gender of the user, positive for females and negative for males. Table 2 shows the distributions of both traits in the two collections.

[6]The cut-off age was determined for maintaining a balance between both groups of users, especially in the Reddit collection.

Table 2 Distribution of the type of users and the polarity of their posts in the Reddit and Twitter collections, according to their automatically inferred labels

Partition	Trait	Reddit				Twitter			
		Category		Polarity (%)		Category		Polarity (%)	
		Dep	NDep	Pos	Neg	Dep	NDep	Pos	Neg
Training	Male	69	630	74%	26%	1216	2705	55%	45%
	Female	66	122	84%	16%	1410	2668	58%	41%
	Young	80	324	70%	30%	1153	2105	66%	34%
	Senior	55	428	72%	28%	1473	3268	71%	29%
Test	Male	42	610	70%	30%				
	Female	37	131	88%	12%				
	Young	47	263	63%	37%				
	Senior	32	478	61%	39%				

Table 3 F1 results over the positive class using the proposed approach, which combines specialized classifiers per users' traits and the BoP representation

Approach	Representation	Reddit	Twitter
Single Classifier	BoW	0.63 (± 0.04)	0.85 (± 0.01)
	BoP	0.64 (± 0.04)	0.85 (± 0.01)
Age-based classifiers	BoW	0.65 (± 0.07)	0.86 (± 0.03)
	BoP	0.68 (± 0.01)	0.88 (± 0.05)
Gender-based classifiers	BoW	0.64 (± 0.04)	0.87 (± 0.04)
	BoP	**0.71** (± **0.02**)	**0.89** (± **0.01**)

They indicate that the number of male and female users, as well as young and senior, are similar in the Twitter collection, however, they strongly differ in Reddit, where men and senior users are more abundant than women and young users, respectively.

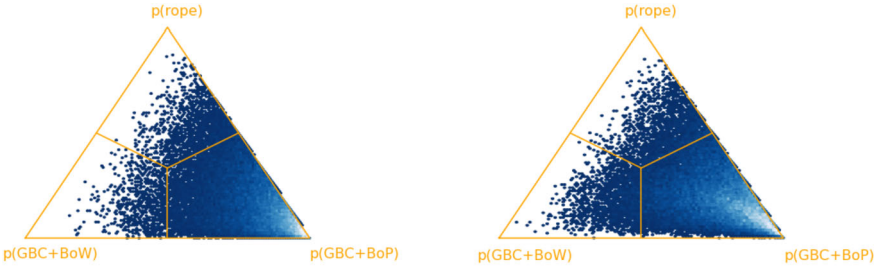
6 Results

6.1 Overall performance evaluation

The main goal of the experiments carried out was to evaluate the suitability of the proposed approach for the depression detection task, considering data from two different social media. Indirectly, our goal was to determine the relevance of using specialized classifiers for the different kinds of users as well as of taking into account the polarity of the context of occurrence of the words. Table 3 shows the results of the different configurations we considered. It includes results of classifiers that were specialized according to the age and gender of the users, referred as *age-based classifiers* and *gender-based classifiers*, respectively. It also shows the results for the standard approach that uses one single classifier for all users (named as *single classifier*). For each of these configurations, it presents results with the traditional BoW representation, as well as with the proposed BoP representation, which provides sentiment awareness.

From the results of Table 3, the following can be highlighted:

First, the proposed approach, which jointly uses the words, the polarities of their posts, and the profile traits of the users to distinguish those suffering from depression, *clearly outperformed the baseline approach* that corresponds to the use of one single classifier for all users with a BoW representation. This is especially evident in the Reddit collection, where the best result of the proposed approach shows a difference around 8% with respect to the baseline; for Twitter the difference was around 4%.



(a) **Reddit**:
$P(\text{GBC+BoW} > \text{GBC+BoP}) = 0.01088$
$P(\text{rope}) = 0.00132$
$P(\text{GBC+BoP} > \text{GBC+BoW}) = 0.98780$

(b) **Twitter**:
$P(\text{GBC+BoW} > \text{GBC+BoP}) = 0.04902$
$P(\text{rope}) = 0.00272$
$P(\text{GBC+BoP} > \text{GBC+BoW}) = 0.94826$

Figure 3 Results of the Bayesian Hierarchical test comparing the gender-based classifiers (GBC) using the BoP and BoW representations

Second, neither of the two components of the proposed approach (namely, specialized classifiers per users' traits and the BoP representation), is by itself good enough to obtain significantly better results than the baseline; *they have to be used in combination*. This observation is supported by the small differences between the results of the single classifier using BoW and BoP, as well as by the differences between the trait-based classifiers and the single classifier when using BoW as representation. In general, these results corroborate our initial intuition that the words' discriminative value for the detection of depression depends on both, the characteristics of the person who is writing, and the polarity of the messages where they occur.

Third, the age-based classifier with the BoP representation shows a good performance (better than the baseline); however, the best results were achieved when considering the *gender-based classifier*, which reached outstanding results in both collections, 0.71 and 0.89 for Reddit and Twitter, respectively. Despite this difference, we believe our results are not entirely conclusive regarding the advantage of using classifiers by gender instead of by age, since they depend on the method used to infer these attributes and, on the other hand, age-based classifiers could be improved when considering finer age groups.

To investigate whether the differences of the results between the approaches were statistically significant, we performed a statistical analysis. According to the t-test, and using a $p = 0.05$, the results from the trait-based classifiers with the BoP representation are significantly better than the results from the baseline approach, and also than the results from these classifiers but with the BoW representation. To deepen the comparison, we applied the Non-Parametric Bayesian Hierarchical test [53] over the results of the gender-based classifiers using BoP and BoW. Figure 3 shows the results of this analysis; they indicate a greater probability that the gender-based classifiers obtained better results using BoP than BoW.

6.2 Comparison with state-of-the-art results

Table 4 compares the best results of the proposed approach against the best results previously reported for the two used collections. These results correspond to the following works: [30], which achieved the best performance in the eRisk 2018 shared task, and employed user-level linguistic metadata, a bag of words representation, neural word embed-

Table 4 Comparison of the performance of our gender-based classifiers with the BoP representation against state-of-the-art results. The results marked with an "*" correspond to approaches using both text and image features

Dataset	Approach	F1 (positive class)
Reddit	**Gender-based classifiers using BOP**	**0.71** (±–**0.02**)
	BCSGB model [30]	0.64
	gSC model [54]	0.59
	BoSE-ngrams [11]	0.63
	DeepBoSE [55]	0.65
		Macro-F1
Twitter	**Gender-based classifiers using BOP**	**0.87** (± **0.01**)
	MDL [50]	0.85*
	DFC+FC [56]	78.5 (± 1.2)*
	GRU+VGG-NET+COMMA [57]	**0.90***
	BiGRU-CNN [58]	0.85

dings from Glove, and a convolutional neural network; [54], in which the task was modeled as a one-class classification problem in order to deal with the uncertainty regarding negative instances; [50] and [56], in which text and image features were used, as well as the posting behavior of users; [58], which used word embeddings and a combination of recurrent and convolutional neural networks; [57], which proposed a multimodal approach, using text and images, and applying a complex deep neural network architecture; [11], which proposed to model the users by frequency histograms of fine-grained emotions; and [55], which extended the previous work by including the learning of the fine-grained emotions into an end-to-end architecture.

In general, the results from Table 4 indicate that the proposed approach, despite its simplicity (it is based on a word-based representation), performs better than most state-of-the-art methods based on complex deep learning models. For example, in the Reddit collection, it had a gain of 7% with respect to the best previous result, whereas, in the Twitter corpus, it performed better than the works using only textual features, but showed a slightly smaller F1 result (3% less) than a multimodal approach using both text and image information.

6.3 Assessing the robustness of the proposed approach to different classifiers

The previous sections showed results of the proposed approach when it was used in conjunction with a Bagging of Decision Trees as classifier, highlighting its relevance for the depression detection task. The goal of this section is to study the robustness of the approach when other machine learning algorithms are used to build the classifier. More precisely, we aim to assess the ability of the approach to achieve a similar performance when used in conjunction with other classifiers.

For this experiment, we consider two classifiers commonly used in text classification tasks, namely, a Support Vector Machine (SVM) and a Random Forest (RF). Table 5 shows the results obtained with both classifiers in the Reddit and Twitter collections. Analyzing the results under the approximation of a "single classifier", we can observe that BoP is better than BoW, regardless of the classifier used, therefore confirming the relevance of taking into account the polarities of the posts for the depression detection task. On the other hand, the trait-based classifiers achieved in all cases better results than the single classifiers, reaffirming the adequacy of specializing the classification process according to the different kinds of users. In general, these experiments show that the performance of the

Table 5 Performance of the proposed approach using different classifiers. BoW combining specialized classifiers per users' traits is showed as a baseline

Approach	Rep.	Reddit		Twitter	
		SVM	RF	SVM	RF
Single Classifier	BoW	0.57	0.54 (± 0.05)	**0.85** (± **0.02**)	0.75 (± 0.03)
	BoP	**0.59**	**0.60** (± **0.03**)	**0.85** (± **0.01**)	**0.78** (± **0.04**)
Age-based classifiers	BoW	0.58	0.60 (± 0.07)	0.82 (± 0.02)	0.78 (± 0.04)
	BoP	**0.62**	**0.64** (± **0.06**)	**0.85** (± **0.02**)	**0.85** (± **0.01**)
Gender-based classifiers	BoW	0.60	0.58 (± 0.08)	**0.86** (± **0.05**)	0.85 (± 0.02)
	BoP	**0.65**	**0.66** (± **0.03**)	0.85 (± 0.00)	**0.86** (± **0.03**)

proposed approach is robust to the selection of the classifier, and that it mostly depends on the modeling of the words, the polarities of their posts, and the profile traits of the users. Nonetheless, as shown in Table 3, the best results were obtained when using the Bagging of Decision Trees as classifier.

7 Analysis and discussion of results

7.1 Users' traits, posts' polarities, and the discriminative value of words

The previous experiments confirmed that words alone are not enough to fully distinguish between social media users who suffer from depression and those who do not. Furthermore, they highlighted the relevance of considering some characteristics of the person who is writing, and the polarity of the messages where they occur, to improve the detection of those suffering from depression. In order to understand the influence of these two aspects on the discriminative value of words, we analyzed the *information gain* (IG) of the features from the BoP representation, for our two trait-based classifiers. Interestingly, we observed that some words present an important difference in their IG values when used in contexts of different polarity. Figure 4 exemplifies some of these words, indicating the relative importance of their IG values in positive (blue bars) and negative (red bars) contexts, and for the gender and age-based classifiers. For example, for discriminating between male depressed and non-depressed users in Reddit, terms related to relationships (e.g., *friendship* and *dating*) were highly relevant when they occurred in positive contexts, whereas words such as *family* and *sexual* were more informative when occurring in negative contexts. In the case of female users, words such as *attractive, sex*, and *family* were more important when occurring in positive than in negative contexts, and words such as *boyfriend* were good discriminators when they occurred in negative contexts. It was interesting to notice that some words were relevant for both gender-based classifiers, but with a presence in distinct contexts, that was the case of *family*, which was highly discriminative in negative posts for males and in positive posts for females. Table 6 shows some example posts with this word. Also in Reddit, words like *dating* and *boyfriend* in positive posts, and *virgin*, *weight* and *relationship* in negative posts, were relevant to distinguish between depressed and non-depressed young users. Interestingly, the word *relationships* was also relevant for the case of senior users, but when it occurred in positive posts. Table 6 also contrasts some example posts from young and senior users containing this word.

In the case of Twitter things were not very different. Posts from depressed and non-depressed users were distinguished by interests and concerns usually associated to their gender and age group. For example, we found a high relevance of the word *calories* when used in negative contexts to identify users, mainly young women, who suffer from depres-

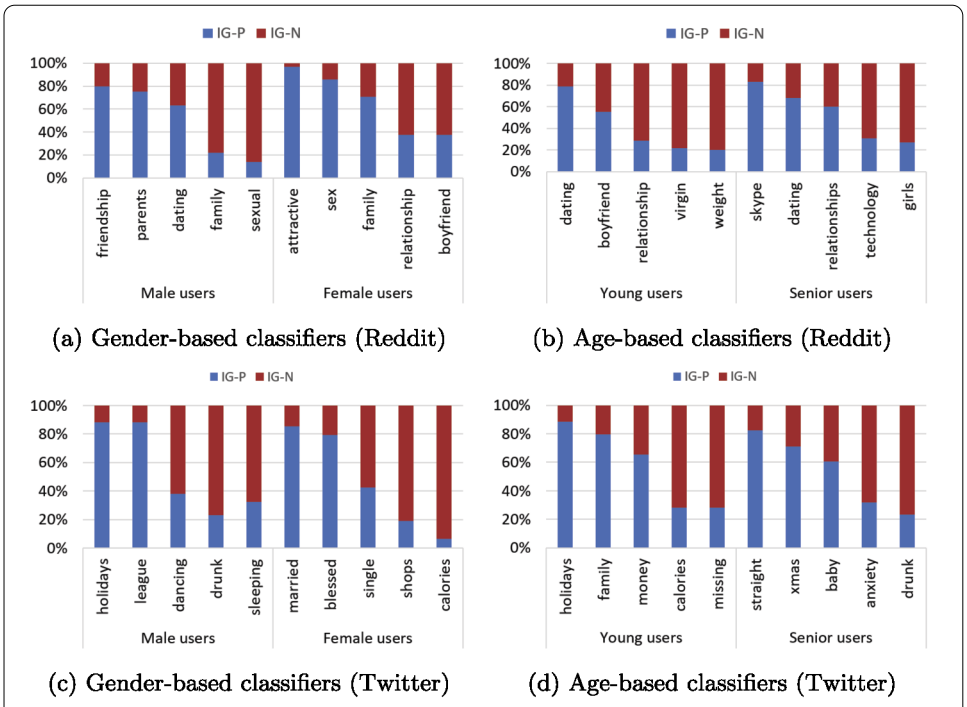


Figure 4 Examples of words that show contrasting information gain values depending on the users traits and polarity of their occurrences. Each bar corresponds to a word and indicates the relative importance of its IG values in positive (blue part) and negative (red part) posts

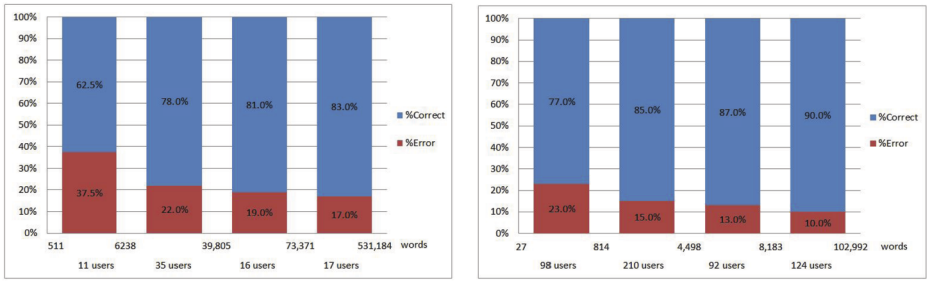
Table 6 Example posts of two discriminative words, which show the relevance of considering contextual information (users' traits and occurrences' polarity) for a proper interpretation

Posts containing the word *family*	Posts containing the word *relationship(s)*
Male users, negative posts	**Senior users, positive posts**
*My living expenses are increasing by 10%... I eat shit food, feed my **family** the best I can, get into arguments with the warden about leaving lights on and the doors open when its 110 degrees outside*	*Maybe you can think of it as going out with a friend rather than a date. After all most good **relationships** between couples are actually just good friendships that eventually go a stage further.*
*Rent! Can't live in the ghetto! Got a **family** to protect! That's almost a grand with even more insurance and fees up my ass.*	*Try not to let your past experiences define your future **relationships**. There are lots of people out there who are bad for us, but there are also some who are good.*
Female users, positive posts	**Young users, negative posts**
*I do owe it to this **family** to be the best person I can be. They are so nice to me. So caring and they trust me a lot. I just want to be a pleasant person to be around, for them.*	*My boyfriend has got borderline personality disorder...it is hard to describe the wicked tangles of guilt and sometimes abuse that such a **relationship** can mean.*
*The **family** you make is more important than the family you come from. Something I live by.*	*I saw posts on Facebook that she's in a **relationship** with a guy I know and hate, and they were chatting back and forth and I've been crying ever since.*

sion. Similarly, the presence of the word *drunk* in negative contexts showed high relevance to detect senior male users who suffer from depression.

7.2 Error analysis: on the effect of the users' history length

As previously mentioned, the detection of depression is commonly handled as a text classification problem, and, therefore, the length of the users' post histories can have some



(a) Reddit (b) Twitter

Figure 5 Percentages of classification errors by users' history lengths (number of posts), in the Reddit and Twitter collections. The statistics correspond only to users from the positive class (i.e., depressed users)

impact on the performance. We carried out an error analysis in the detection of depressed users according to this variable. Figure 5 presents the results from this analysis, considering four length ranges,[7] and indicating the percentages of correct and incorrect predictions.

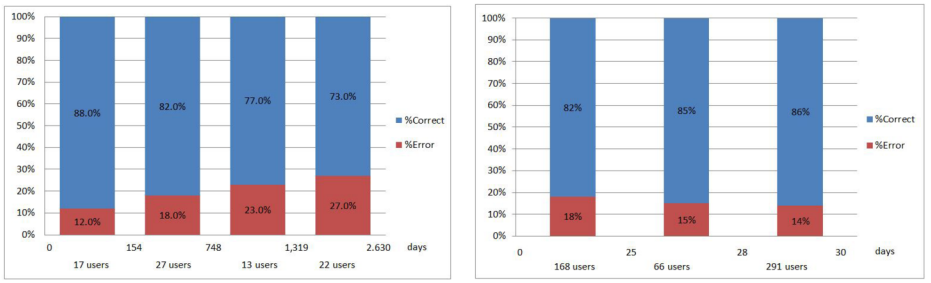
From Fig. 5, it can be observed that in both collections, the higher the number of posts by the users, the lower the prediction errors. For example, the error percentage is only around 17% and 10% for the largest post histories, while it is around 37.5% and 23% for the smallest, in Reddit and Twitter respectively. These results clearly indicate that the more evidence is available on a user, the greater the confidence on the decision of whether or not he/she suffers from depression.

7.3 Error analysis: on the effect of the temporal expansion of users' histories

We also analyzed the error percentages with respect to the temporal expansion of the users' post histories, defining it as the number of days between the first and last post from a user. For this analysis, we considered four and three intervals of time[8] for the Reddit and Twitter collections, respectively. Figure 6 shows the results of this analysis. In the Reddit collection, the error percentage increased as the temporal expansion increased. In particular, it exceeded 20% when the users were observed for very long periods (more than two years). This result may be due to the fact that during a long period users could receive treatment or even come out of their depression. On the other hand, in the case of Twitter, the error percentage decreased when users were observed for more time. Although these results may seem contradictory, they are actually complementary; it is important to highlight the difference in the size of the observation intervals, because in the case of the Twitter collection, no user shows a temporal expansion greater than 30 days. Hence, as an integrated conclusion, we can say that it is expected to have a reduction in the detection errors when more user information is analyzed; however, if users are observed for long periods, there is a greater chance for depressive traits to fade or blur as a result of various factors, such as, for example, a successful treatment.

[7] These ranges were determined by assuming a normal distribution; their limits correspond to the media, and the media ± a standard deviation.

[8] Similar to the previous analysis, we determined the intervals by assuming a normal distribution, and establishing the interval boundaries in the media, and the media ± a standard deviation. However, in the case of Twitter, we decided to merge the third and fourth intervals due to the small number of instances that fall in the last one.



(a) Reddit (b) Twitter

Figure 6 Error percentages by temporal expansion in the Reddit and Twitter collections. The x-axis represents the observation time intervals per user; in Reddit the maximum temporal expansion is of 2630 days, in Twitter only 30 days

8 Conclusions and future work

In this work we proposed a novel *profile-based sentiment-aware approach* for the detection of depression in social media. The main idea behind this approach is to leverage implicit information of the words' contexts to enhance the detection of signs of depression, under the assumption that the words' discriminative value for this task depends on the *characteristics of the user* who is writing and on the *polarity of the messages* where they occur.

Through the paper, we presented an experimental study that evaluates the proposed approach in two social media collections, one from Reddit and another from Twitter, which considerably differ in the number of users, the length of their posts, and the temporal expansion of their histories. In general, the results achieved were encouraging, since they outperformed those of state-of-the-art corresponding to computationally more expensive methods, indicating the suitability of the approach for the task at hand. In particular, we obtained the following conclusions: (*i*) considering specific classifiers based on the *users' gender* is more useful than based on their age, which could indicate that the differences in the expression of depressed people are more noticeable among men and women, than among young people and adults. However, as we previously pointed out, this observation could be changed when considering finer age groups. (*ii*) the discriminative value of words as depression markers greatly depends on their contexts' *polarity*, positive or negative, but at the same time, the adequate interpretation of it varies according to the type of user, therefore, trait-based classifiers and the dual word-sentiment representation need to be used in combination. (*iii*) factors such as the length and temporal expansion of the users' post histories had an important influence on the approach performance, particularly, we observed *better results when more user information was analyzed*; however, when users were observed for very long periods the performance tended to decrease.

As future work, we plan to combine the proposed approach with different text representations, such as those based on word embeddings or LIWC. We also consider to explore other author profiling methods to infer the gender and age of users, as well as to use finer age groups. Likewise, we have in mind to apply attention models under a trait-based approach to automatically learn relevant word-sequences depression patterns. Finally, the good results achieved also motive us to carry out a similar study in other related tasks, such as, for example, in the detection of aggressive comments in social media, this under

the premise that different types of users tend to show different types of aggressiveness, and that the posts' polarity is a key issue on the correct interpretation of bad words and other idiomatic expressions.

Acknowledgements
This research was partially supported by INAOE and CONACYT-Mexico (scholarship CVU-905721 and grant CB-2015-01-257383).

Funding
This publication has received the support of CONACyT and INAOE, Mexico.

Availability of data and materials
In this work two datasets were used, one from the eRisk 2018 forum for depression detection task[9] and the collection of Twitter users described in [50].

Declarations

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
JJT performed the experimentation. RMO carried out the analysis of the use of profiling and sentiment information. MM and LV supervised the research and contributed to the analyzes of results and errors. All authors wrote, reviewed, and approved the final manuscript.

Author details
[1]Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Luis Enrique Erro #1, Sta María Tonantzintla, San Andrés Cholula, Puebla, México. [2]Universidad Politécnica de Tulancingo, Calle Ingenierías #100, Huapalcalco, Tulancingo de Bravo, Hidalgo, México.

Publisher's Note

Received: 29 April 2021 Accepted: 24 October 2021 Published online: 06 November 2021

References

1. Patton G, Coffey C, Sawyer S, Viner R, Haller D, Bose K, Vos T, Ferguson J, Mathers C (2009) Global patterns of mortality in young people: a systematic analysis of population health data. Lancet 374:881–892. https://doi.org/10.1016/s0140-6736(09)60741-8
2. Losada DE, Crestani F (2016) A test collection for research on depression and language use. In: Fuhr N, Quaresma P, Gonçalves T, Larsen B, Balog K, Macdonald C, Cappellato L, Ferro N (eds) Experimental IR meets multilinguality, multimodality, and interaction. Springer, Cham, pp 28–39. https://doi.org/10.1007/978-3-319-44564-9_3
3. Piccinelli M, Wilkinson G (2000) Gender differences in depression. Critical review. Br J Psychiatry 177:486–492. https://doi.org/10.1192/bjp.177.6.486
4. Rude S, Gortner E-M, Pennebaker J (2004) Language use of depressed and depression-vulnerable college students. Cogn Emot 18(8):1121–1133. https://doi.org/10.1080/02699930441000030
5. Losada DE, Crestani F, Parapar J (2018) Overview of erisk: early risk prediction on the Internet. In: Bellot P, Trabelsi C, Mothe J, Murtagh F, Nie JY, Soulier L, SanJuan E, Cappellato L, Ferro N (eds) Experimental IR meets multilinguality, multimodality, and interaction. Springer, Cham, pp 343–361. https://doi.org/10.1007/978-3-319-98932-7_30
6. Losada DE, Crestani F, Parapar J (2017) Clef 2017 erisk overview: early risk prediction on the Internet: experimental foundations. In: Proceedings of the 8th international conference of the CLEF association, CLEF 2017, Dublin, Ireland. http://ceur-ws.org/Vol-1866/invited_paper_5.pdf
7. Coppersmith G, Dredze M, Harman C, Hollingshead K, Mitchell M (2015) CLPsych 2015 shared task: depression and PTSD on Twitter. In: Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality. Association for Computational Linguistics, Denver, pp 31–39. https://doi.org/10.3115/v1/W15-1204
8. Yates A, Cohan A, Goharian N (2017) Depression and self-harm risk assessment in online forums. In: Proceedings of the 2017 conference on empirical methods in natural language processing. Association for Computational Linguistics, Copenhagen, pp 2968–2978. https://doi.org/10.18653/v1/D17-1322
9. Trotzek M, Koitka S, Friedrich CM (2020) Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. IEEE Trans Knowl Data Eng 32(3):588–601. https://doi.org/10.1109/tkde.2018.2885515

[9]https://erisk.irlab.org/

10. Chen X, Sykora MD, Jackson TW, Elayan S (2018) What about mood swings: identifying depression on Twitter with temporal measures of emotions. In: Companion proceedings of the web conference 2018. WWW '18. International world wide web conferences steering committee, republic and canton of Geneva, CHE, pp 1653–1660. https://doi.org/10.1145/3184558.3191624
11. Aragón ME, López-Monroy AP, González-Gurrola LC, Montes-y-Gómez M (2019) Detecting depression in social media using fine-grained emotions. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). Association for Computational Linguistics, Minneapolis, pp 1481–1486. https://doi.org/10.18653/v1/N19-1151
12. Wolohan J, Hiraga M, Mukherjee A, Sayyed ZA, Millard M (2018) Detecting linguistic traces of depression in topic-restricted text: attending to self-stigmatized depression with NLP. In: Proceedings of the first international workshop on language cognition and computational models. Association for Computational Linguistics, Santa Fe, pp 11–21. https://aclanthology.org/W18-4102
13. Resnik P, Garron A, Resnik R (2013) Using topic modeling to improve prediction of neuroticism and depression in college students. In: Proceedings of the 2013 conference on empirical methods in natural language processing. Association for Computational Linguistics, Seattle, pp 1348–1353. https://aclanthology.org/D13-1133
14. Burdisso SG, Errecalde M, Montes-y-Gómez M (2019) A text classification framework for simple and effective early depression detection over social media streams. Expert Syst Appl 133:182–197. https://doi.org/10.1016/j.eswa.2019.05.023
15. Girgus JS, Yang K, Ferri CV (2017) The gender difference in depression: are elderly women at greater risk for depression than elderly men? Geriatr (Basel) 2(4):35. https://doi.org/10.3390/geriatrics2040035
16. Sloan DM, Sandt AR (2006) Gender differences in depression. J Women's Health 2(3):425–434. https://doi.org/10.2217/17455057.2.3.425
17. Tophoven S, du Prel J-B, Peter R, Kretschmer V (2015) Working in gender-dominated occupations and depressive symptoms: findings from the two age cohorts of the lidA study. J Labour Market Res 48(3):247–262. https://doi.org/10.1007/s12651-014-0165-2
18. Albert PR (2015) Why is depression more prevalent in women? J Psychiatry Neurosci 40(4):219–221. https://doi.org/10.1503/jpn.150205
19. Dejonckheere E, Mestdagh M, Houben M, Erbas Y, Pe M, Koval P, Brose A, Bastian B, Kuppens P (2018) The bipolarity of affect and depressive symptoms. J Pers Soc Psychol 114(2):323–341. https://doi.org/10.1037/pspp0000186
20. Dejonckheere E, Kalokerinos EK, Bastian B, Kuppens P (2019) Poor emotion regulation ability mediates the link between depressive symptoms and affective bipolarity. Cogn Emot 33(5):1076–1083. https://doi.org/10.1080/02699931.2018.1524747
21. Liu LS, Huh J, Neogi T, Quinn KI, Pratt W (2013) Health vlogger-viewer interaction in chronic illness management. In: Proceedings of the SIGCHI conference on human factors in computing systems. CHI conference 2013, pp 49–58. https://doi.org/10.1145/2470654.2470663
22. Mowery DL, Park YA, Bryan C, Conway M (2016) Towards automatically classifying depressive symptoms from Twitter data for population health. In: Proceedings of the workshop on computational modeling of people's opinions, personality, and emotions in social media (PEOPLES), pp 182–191. https://aclanthology.org/W16-4320
23. Hiraga M (2017) Predicting depression for Japanese blog text. In: Proceedings of ACL 2017, student research workshop. Association for Computational Linguistics, Vancouver, pp 107–113. https://doi.org/10.18653/v1/p17-3018
24. Jamil Z, Inkpen D, Buddhitha P, White K (2017) Monitoring tweets for depression to detect at-risk users. In: Proceedings of the fourth workshop on computational linguistics and clinical psychology, pp 32–40. https://doi.org/10.18653/v1/w17-3104
25. Sadeque F, Pedersen T, Solorio T, Shrestha P, Rey-Villamizar N, Bethard S (2016) Why do they leave: modeling participation in online depression forums. In: Proceedings of the fourth international workshop on natural language processing for social media. Association for Computational Linguistics, Austin, pp 14–19. https://doi.org/10.18653/v1/W16-6203
26. Schwartz HA, Eichstaedt J, Kern ML, Park G, Sap M, Stillwell D, Kosinski M, Ungar L (2014) Towards assessing changes in degree of depression through Facebook. In: Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality. Association for Computational Linguistics, Baltimore, pp 118–125. https://doi.org/10.3115/v1/W14-3214
27. Resnik P, Armstrong W, Claudino L, Nguyen T (2015) The University of Maryland CLPsych 2015 shared task system. In: Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality. Association for Computational Linguistics, Denver, pp 54–60. https://doi.org/10.3115/v1/W15-1207
28. Loveys K, Torrez J, Fine A, Moriarty G, Coppersmith G (2018) Cross-cultural differences in language markers of depression online. In: Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic. Association for Computational Linguistics, New Orleans, pp 78–87. https://doi.org/10.18653/v1/W18-0608
29. Funez DG, Ucelay MJG, Villegas MP, Burdisso S, Cagnina LC, Montes-y-Gómez M, Errecalde M (2018) Unsl's participation at erisk 2018 lab. In: Cappellato L, Ferro N, Nie J, Soulier L (eds) Working notes of CLEF 2018—conference and labs of the evaluation forum, Avignon, France, September 10–14, 2018. CEUR workshop proceedings, vol 2125. CEUR-WS.org, Avignon. http://ceur-ws.org/Vol-2125/paper_137.pdf
30. Trotzek M, Koitka S, Friedrich CM (2018) Word embeddings and linguistic metadata at the clef 2018 tasks for early detection of depression and anorexia. In: Cappellato L, Ferro N, Nie J-Y, Soulier L (eds) Working notes of CLEF 2018—conference and labs of the evaluation forum, Avignon, France, September 10–14, 2018 CEUR workshop proceedings, vol 2125. CEUR-WS.org, Avignon. http://ceur-ws.org/Vol-2125/paper_68.pdf
31. Chancellor S, De Choudhury M (2020) Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med 3(1):43. https://doi.org/10.1038/s41746-020-0233-7
32. Khumalo IP, Temane QM, Wissing MP (2012) Socio-demographic variables, general psychological well-being and the mental health continuum in an African context. Soc Indic Res 105(3):419–442. https://doi.org/10.1007/s11205-010-9777-2

33. Preoţiuc-Pietro D, Eichstaedt J, Park G, Sap M, Smith L, Tobolsky V, Schwartz HA, Ungar L (2015) The role of personality, age, and gender in tweeting about mental illness. In: Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality. Association for Computational Linguistics, Denver, pp 21–30. https://doi.org/10.3115/v1/W15-1203
34. Eo A, Ak N, Okechukwu C, Onoh I, Nt O (2020) Socio-demographic characteristics and other factors associated with depressive illness among medical students at the University of Port Harcourt. Insights Depress Anxiety 4(1):040. https://doi.org/10.29328/journal.ida.1001018
35. Angst J, Gamma A, Gastpar M, Lépine J-P, Mendlewicz J, Tylee A (2002) Gender differences in depression. Eur Arch Psychiatry Clin Neurosci 252(5):201–209. https://doi.org/10.1007/s00406-002-0381-6
36. Nolen-Hoeksema S (2001) Gender differences in depression. Curr Dir Psychol Sci 10(5):173–176. https://doi.org/10.1111/1467-8721.00142
37. Call JB, Shafer KM (2018) Gendered manifestations of depression and help seeking among men. Am J Men's Health 12:41–51. https://doi.org/10.1177/1557988315623993
38. McCrae N, Gettings S, Purssell E (2017) Social media and depressive symptoms in childhood and adolescence: a systematic review. Adolesc Res Rev 2:315–330. https://doi.org/10.1007/s40894-017-0053-4
39. Frenzel A (2014) Teacher emotions. In: Linnenbrink-Garcia EA, Pekrun R (eds) International handbook of emotions in education. Routledge, New York, pp 494–519
40. Munezero M, Montero CS, Sutinen E, Pajunen J (2014) Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text. IEEE Trans Affect Comput 5(2):101–111. https://doi.org/10.1109/TAFFC.2014.2317187
41. Rottenberg J (2017) Emotions in depression: what do we really know? Annu Rev Clin Psychol 13:241–263. https://doi.org/10.1146/annurev-clinpsy-032816-045252
42. Blanco I, Joormann J (2017) Examining facets of depression and social anxiety: the relation among lack of positive affect, negative cognitions, and emotion dysregulation. Span J Psychol. https://doi.org/10.1017/sjp.2017.43
43. Forbes EE, Williamson DE, Ryan ND, Dahl RE (2004) Positive and negative affect in depression: influence of sex and puberty. Ann NY Acad Sci 1021(1):341–347. https://doi.org/10.1196/annals.1308.042
44. Bathina KC, Ten Thij M, Lorenzo-Luaces L, Rutter LA, Bollen J (2021) Individuals with depression express more distorted thinking on social media. Nat Hum Behav 5(4):458–466. https://doi.org/10.1038/s41562-021-01050-7
45. Wang Y, Hou Y, Che W, Liu T (2020) From static to dynamic word representations: a survey. Int J Mach Learn Cybern 11(7):1611–1630. https://doi.org/10.1007/s13042-020-01069-8
46. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations. In: Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long papers). Association for Computational Linguistics, New Orleans, pp 2227–2237. https://doi.org/10.18653/v1/N18-1202
47. Qadir A, Riloff E, Walker M (2015) Learning to recognize affective polarity in similes. In: Proceedings of the 2015 conference on empirical methods in natural language processing. Association for Computational Linguistics, Lisbon, pp 190–200. https://doi.org/10.18653/v1/D15-1019
48. Hassan AU, Hussain J, Hussain M, Sadiq M, Lee S (2017) Sentiment analysis of social networking sites (sns) data using machine learning approach for the measurement of depression. In: 2017 international conference on information and communication technology convergence (ICTC), pp 138–140. https://doi.org/10.1109/ICTC.2017.8190959
49. Tian L, Lai C, Moore J (2018) Polarity and intensity: the two aspects of sentiment analysis. In: Proceedings of the first grand challenge and workshop on human multimodal language (challenge-HML), pp 40–47. https://doi.org/10.18653/v1/W18-3306
50. Shen G, Jia J, Nie L, Feng F, Zhang C, Hu T, Chua T-S, Zhu W (2017) Depression detection via harvesting social media: a multimodal dictionary learning solution. In: Proceedings of the twenty-sixth international joint conference on artificial intelligence, IJCAI-17, pp 3838–3844. https://doi.org/10.24963/ijcai.2017/536
51. Baccianella S, Esuli A, Sebastiani F (2010) SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the seventh international conference on language resources and evaluation (LREC'10). European Language Resources Association (ELRA), Valletta. http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf
52. Sap M, Park G, Eichstaedt J, Kern M, Stillwell D, Kosinski M, Ungar L, Schwartz HA (2014) Developing age and gender predictive lexica over social media. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). Association for Computational Linguistics, Doha, pp 1146–1151. https://doi.org/10.3115/v1/D14-1121
53. Corani G, Benavoli A, Demšar J, Mangili F, Zaffalon M (2017) Statistical comparison of classifiers through Bayesian hierarchical modelling. Mach Learn 106(11):1817–1837. https://doi.org/10.1007/s10994-017-5641-9
54. Aguilera J, Hernández Farías DI, Ortega-Mendoza RM, Montes-y-Gómez M (2021) Depression and anorexia detection in social media as a one-class classification problem. Appl Intell 51(8):6088–6103. https://doi.org/10.1007/s10489-020-02131-2
55. Lara JS, Aragon ME, Gonzalez FA, Montes-y-Gómez M (2021) Deep bag-of-sub-emotions for depression detection in social media. arXiv preprint 2103.01334
56. Shen T, Jia J, Shen G, Feng F, He X, Luan H, Tang J, Tiropanis T, Chua T-S, Hall W (2018) Cross-domain depression detection via harvesting social media. In: Proceedings of the twenty-seventh international joint conference on artificial intelligence, IJCAI-18. International Joint Conferences on Artificial Intelligence Organization, Stockholm, pp 1611–1617. https://doi.org/10.24963/ijcai.2018/223
57. Gui T, Zhu L, Zhang Q, Peng M, Zhou X, Ding K, Chen Z (2019) Cooperative multimodal approach to depression detection in Twitter. In: Proceedings of the AAAI conference on artificial intelligence, vol 33. pp 110–117. https://doi.org/10.1609/aaai.v33i01.3301110
58. Zogan H, Razzak I, Wang X, Jameel S, Xu G (2020) Explainable depression detection with multi-modalities using a hybrid deep learning model on social media. arXiv preprint 2007.02847