Temporal Mood Variation: at the CLEF eRisk-2018 Tasks for Early Risk Detection on The Internet

Waleed Ragheb, Bilel Moulahi, Jérôme Azé, Sandra Bringay, Maximilien Servajean

Temporal Mood Variation: at the CLEF eRisk-2018 Tasks for Early Risk Detection on The Internet

Waleed Ragheb[1,2], Bilel Moulahi[1,2], Jérôme Azé[1,2], Sandra Bringay[1,3], and Maximilien Servajean[1,3]

[1] LIRMM UMR 5506, CNRS, University of Montpellier, Montpellier, France
[2] IUT de Béziers, University of Montpellier, Béziers, France
[3] AMIS, Paule Valery University - Montpellier 3 , Montpellier, France
{waleed.ragheb,bilel.moulahi,jerome.aze,sandra.bringay,maximilien.servajean}@lirmm.fr

Abstract. Two tasks are proposed at CLEF eRisk-2018 on predicting mental disorder using Users posts on Reddit. Depression and anorexia disorders are considered to be detected as early as possible. In this paper we present the participation of LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier) in both tasks. The proposed architectures and models use only text information without any hand-crafted features or dictionaries to model the temporal mood variation detected from users posts. The proposed models use two learning phases through exploration of state-of-the-art text vectorization. The proposed models perform comparably to other contributions while experiments shows that document-level outperformed word-level vectorizations.

Keywords: Classification, Word2vec, Doc2vec, Temporal Variation, MLP, Depression, Anorexia

1 Introduction

Depression is a common mental disorder. Globally, more than 300 million people of all age stages suffer from depression [13]. It has a direct and indirect effect on the economic growth because of its major impact on the productivity. Depression also has dramatic consequences not only for those affected but also for their families and their social and work related environments [27]. It may be the psycho-physiological basis for panic and anxiety symptoms. Panic disorder has been increasingly focused on health services and the media, where it affects young people aged 20-40. The incidence of these disorders affects 22% of the adult world population. At its worst consequences, depression is one of the major causes of suicide [2]. Another common mental disorder is Anorexia which is described as an eating disorder. It is characterized by low weight, worry of gaining weight, and a powerful need to be skinny, leading to food restriction. Many who suffer from eating disorder see themselves as overweight although they could be thin [9].

Individuals with eating disorders have also been shown to have lower employment rates, in addition to an overall loss of earnings. Eating disorder sufferers who are experiencing an overall loss in earnings associated with their illness are also magnified by the excess of health-care costs. According to the National Eating Disorder Association (NEDA), up to 70 million people worldwide suffer from eating disorders [1]. Eating disorder symptoms are beginning earlier in both males and females. As estimated, 1.1 to 4.2 percent of women suffer from anorexia at some point in their lifetime [6]. Young people between the ages of 15 and 24 with anorexia have 10 times the risk of dying compared to their same-aged peers.

Social media is becoming increasingly used not only by adults but also at different age stages. Mental disordered patients also turn to online social media and web forums for information on specific conditions and emotional support [8, 7]. Even though social media can be used as a very helpful tool in changing a person's life, it may cause such conflicts that can have a negative impact. This puts responsibilities for content and community management for monitoring and moderation. With the increasing number of users and their contents these operations turn out to be extremely difficult. Many social media try to deal with this problem by reactive moderation. In reactive moderation, users report any inappropriate, negative or risky user generated contents. However it may reduce the workload or the cost of moderating, it is not enough especially for handling mental disordered user's threads or posts.

Previous researches on social media have established the relationship between an individual's psychological state and his\her linguistic and conversational patterns [25, 11, 23, 14]. This motivate the task organizers to initiate the pilot task for detecting depression from user posts on Reddit[1] in eRisk-2017 [15]. In eRisk-2018 the extension of the study was planned to include detection of anorexia. The main idea is to detect such psychological problems from users posts as early as possible.

In this paper, we present the participation of LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier) in both tasks for early detection of depression and anorexia in eRisk-2018. The originality of our approach is to perform the detection through two main learning phases using text vectorizations. The first phase uses Bayesian rule inversion to construct a time series representing temporal mood variation through users posts. The second phase is to build variable length time series classification model to obtain the proper decision. The main idea is to give a decision once the time series prove clear signs of mental disorder from current and previous mood extracted from the content.

The rest of the paper is organized as follows. In section 2, a short description of these tasks is introduced. Then in Section 3, the related work is introduced. Section 4 describes the training and testing datasets for both tasks. Section 5 presents the experimental setup of the proposed models. In Section 6, the evalu-

[1] Reddit is an open-source platform where community members (red-ditors) can submit content (posts, comments, or direct links), vote submissions, and the content entries are organized by areas of interests (subreddits).

ation results are presented. The conclusions of the experiments and participation are stated in Section 7.

2 Tasks Description

In CLEF eRisk 2018, two tasks are presented [16]. Both tasks are considered as a binary classification problem. The first task is to discriminate between depressed and non-depressed users while the second one is between users diagnosed with anorexia and non-anorexia. The datasets are a dated textual data of user posts and comments -posts without titles- on Reddit. The training data is divided into 10 chunks in chronological order. Each chunk contains 10% of the user's posts. A description of the datasets for both tasks is presented in Section 4. The goal is not only to perform classification but also to do it as soon as possible using minimum amount of data or chunks for each user.

The test datasets also comes with chunks exactly the same way the training data is divided. The tasks organizers give one week for processing each test chunk for both tasks before firing a decision. The decision could be one of the classes or could be postponed for future chunks. At the end of the 10^{th} chunk, all classification propositions must have been submitted. Each team could only participate in one or two tasks and submit at most five runs per task.

For evaluation, the classical classification performance measures (Precision, recall and F1) are computed for each run. In addition error measures called Early Risk Detection Error ($\text{ERDE}_{5,50}$) are computed. It takes into account the correctness of the (binary) decision and the delay taken by the system to make the decision [15].

The training data was released on November 30^{th} 2017. After more than two months and on February 6^{th} the first chunk of test data was released. On a weekly basis a new chunk was being released until the last 10^{th} chunk on April 10^{th}. The evaluation results were announced on 24^{th}April.

3 Related Work

Recent psychological studies showed the correlation between person's mental status and mood variation over time [15, 3, 24, 4]. It is also evident that some mental disordered may have chronic week-to-week mood instability. It is a common presenting symptom for people with a wide variety of mental disorders, with as many as 8 of 10 patients reporting some degree of mood instability during assessment. These studies suggest that clinicians should screen for temporal mood variation across all common mental health disorders [24].

Concerning text representation, traditional Natural Language Processing (NLP) modules starts with extracting some important features from text. These features could be for example the count or frequency of specific words, predefined patterns, Part-of-Speech tagging, etc. These hand-crafted features should be selected carefully and sometimes with an expert view. However these

features are interesting [28, 5], sometimes they loose the sense of generalization. Another recent trend is the use of word and documents vectorization methods. These strategies that convert either words, sentences or even overall documents into vectors take into account all the text not just parts of it. There are many ways to transform a text to high-dimensional space such as term frequency and inverse document frequency (TF-IDF), Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), etc [18]. This direction was revolutionized by Mikolov et al. [20–22] who proposed the Continuous Bag Of Words (CBOW) and skip-gram models known as Word2vec. It is a probabilistic based model that makes use of a two layered neural network architecture to compute the conditional probability of a word given its context. Based on this work Le et al. [12] propose Paragraph Vector model. The algorithm which is also known as Doc2vec learns fixed-length feature representations from variable-length pieces of texts, such as sentences, paragraphs, and documents. Both word vectors and documents vectors are trained using stochastic gradient descent and back-propagation neural network language models. In this paper, we will use both techniques for text representations.

Concerning mood evaluation, one of the interseting work on text distributed representation is the bayesian inversion proposed by Taddy in [26]. It uses Bayes formula to compute the probabilities of a document belonging to a class topic. Given a document d and label y, Bayes formula is:

$$p(y|d) = \frac{p(d|y)p(y)}{p(d)}$$

For classification problems, $p(d)$ can be ignored since d is fixed. $p(d|y)$ is estimated by first training the text vectorization model on a subset of the corpus with label y, then using the skipgram objective composite likelihood as an approximation. As discussed in [26], bayesian inversion will not always outperform other classification methods. It rather provides simple, scalable, interpretable and effective option for classification whenever distributed representations are used. In this paper, we will use bayesian inversion to construct a time series representing temporal mood variation.

4 Datasets

The collection was created as a sequence of XML files, one file per redditor. Each XML file stores the sequence of the redditor's submissions (one entry per submission). Each submission is represented by the submission's title, the submission's text and the submission's date. No other metadata is available [14]. As mentioned before, the redditor's submissions come in ten chunks. Each chunk contains 10% from the overall user submissions. The classification ground truth (golden truth) is given or expected to be predicted for each user. For eRisk-2018 two groups of datasets are provided for each task. A brief description of these datasets is presented in this section.

4.1 Task-1: Depression Dataset

The depression datasets contain submissions from either depressed or non-depressed (controlled) users. The classifications was done manually by the e-Risk organizers. Since the eRisk-2017 pilot task was about the depression classification, in this task two pre-annotated datasets are provided (Training Dataset (2017) and testing DataSets (2017)). Some interesting Statistics and summary for these datasets are provided in Table 1. For eRisk-2018, they provide a test dataset which approximately doubles the number of users from the previous year datasets. However it preserves similar characteristics of the two previous datasets in terms of the number of submissions per users and the length of submissions in sentences or in words. These numbers are just averages but we clearly note the existence of very long/short sentences and submissions. Users' submissions are either posts or comments. Both are described with the same attributes (title, text and date) with empty titles for comments. All datasets collections are unbalanced (more non-depressed users than depressed ones).

Table 1: Summary on Task.1-Depression Datasets

	Training Dataset (2017)	Testing Dataset (2017)	Testing Dataset (2018)
No. of Users (Depressed/Non-Depressed)	486 (83/403)	401(52/349)	820 (79/741)
No. of Submissions	295,023	236,371	544,447
Avg. No. of Submissions/User	607.04	589.45	663.95
No. of Sentences	616,191	541,039	1,336,379
Avg. No. of Sentences per Submission	2.09	2.29	2.45
Avg. Sentence Size (words)	14.34	14.22	14.26
Vocabulary Size	160,599	127,764	222,201

4.2 Task-2: Anorexia Dataset

The Anorexia datasets are similar to the depression datasets in terms of structures and attributes. The classification is done for users diagnosed with anorexia and non-anorexia. eRisk organizers initiated this task for the first time in eRisk-2018. So, there is only one training dataset provided prior to the actual test set. Table 2 describes some statistics on the anorexia datasets. Because of the nature of the task and like depression datasets, the anorexia datasets are also unbalanced.

Table 2: Summary on Task.2-Anorexia Datasets

	Training Dataset (2018)	Testing Dataset (2018)
No. of Users (Anorexia/Non-Anorexia)	152 (20/132)	320 (41/279)
No. of Submissions	84,834	168,507
Avg. No. of Submissions/User	558.12	526.58
No. of Sentences	193,026	370,281
Avg. No. of Sentences per Submission	2.28	2.12
Avg. Sentence Size (words)	14.74	14.30
Vocabulary Size	81,497	103,380

5 Proposed Models

The temporal aspects of the eRisk-2018 tasks inspired us to model the temporal mood variation trough user's text content. The average number of days ranging from the first submission to the last submission is approximately 600 days [14]. So, determining the way in which user's posts and comments vary from positive to negative and vise versa through time is worth inspecting. In the proposed models, time aspects are given as chunks. The main idea is to process user submissions for each chunk and determine the probability of how positive or negative the chunk is. The proposed architecture of our models are shown in Figure. 1.

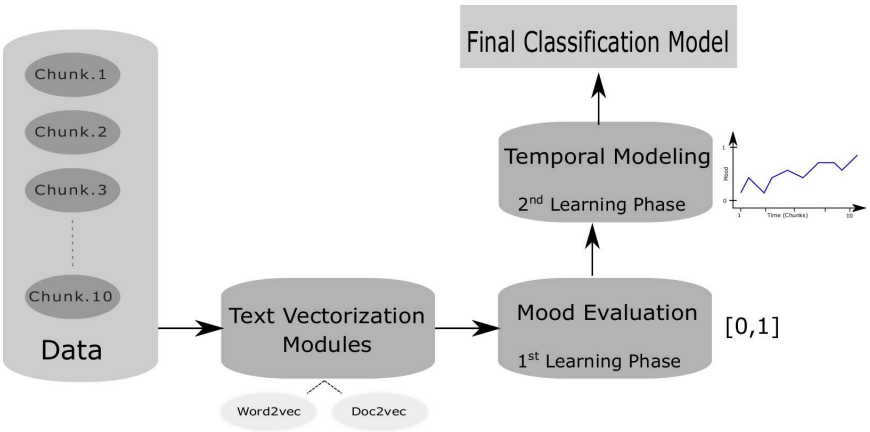

Fig. 1: Block diagram of the main architecture of our models

Step 1 - Text Vectorization Module: The input of this module is the list of textual information divided into ten chunks. The chunks are chronologically ordered as discussed in Section 4. The first step is to build a text vectorization model using all the text chunks. Two state-of-the-art text vectorization models are used. These models are the Word2vec and its evolution, the Doc2vec [20, 12]. The two alternatives of Doc2vec specifications, distributed memory (DM) and distributed bag-of-words (DBOW) are used [12]. We also keep track of the text for each user in every chunk and its label embedded in the model. Also, we built a vectorization model for positive and for negative and did not use any external resources. This module can be considered as an unsupervised learning phase.

Step 2 - Mood Evaluation Module: Our models are based on the work of Matt Taddy in [26] about Bayesian inversion. One of the interesting conclusions from this work is that any distributed representation can be turned into a classifier through inversion via Bayes rule. In our proposed model, we segmented the text of each chunk into sentences and scored each sentence through each vectorization model. The mood of the overall chunk is evaluated simply by normalizing the count of positive and negative sentences using the inversion technique. Each chunk will have a number between [0,1]. This can be considered as the probability of how positive (risky) the chunk is. Processing all chunks leads to a ten-points time series for the ten chunks for each user in the training datasets. Mood evaluation using the inversion technique is considered as the first learning phase in our proposed architecture.

Step 3 - Temporal Modeling Module: Another learning phase is to build machine learning models to learn some patterns from these time series to come up with the final classification model. In the ideal case and for the complete time series, we would have only one model. But since we should not wait for the complete time series we built multiple models for different sizes of time series to be able to give a decision without having to wait for the ten chunks. Figure 2 shows an example of two dimensional representation of the complete time series for the depression task using t-SNE [19]. These time series will be the training set of the second learning phase. The separation between positive and negative users is obvious. It is expected that this separation would not be as ideal as this in testing but it will exists.

Evaluation: For the first task, we used the training set of eRisk-2017 to train the first and second learning models. The testing set of eRisk-2017 was the validation set for our experiments before merging together with eRisk-2017 training set to form the final training set. The evaluation was done for the overall chain of learning phases using the complete ten-points time series. For the second task, we used the Leave-One-Out (LOO) approach in the training data set.

In all the submitted runs for both tasks, only discovered positive users are reported until the end of the last test chunk in which labeling of negative ones was done. The rest of this section will be dedicated to describe the submitted runs for each task. Most of the runs use the same architecture as discussed earlier.

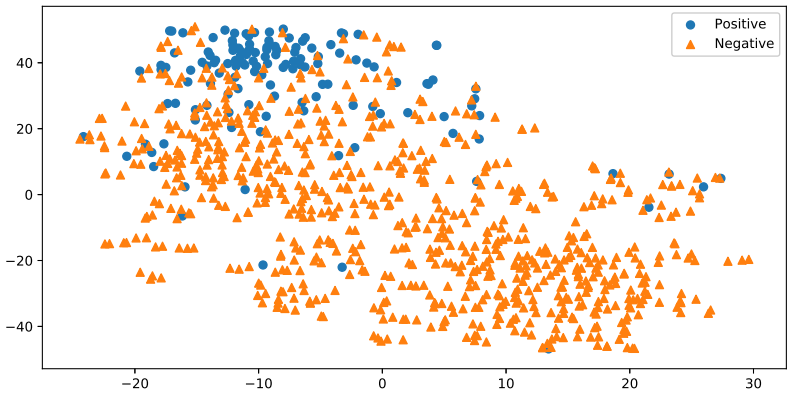

Fig. 2: t-SNE reduced time series information for ten chunks per user in the training datasets in Task-1

5.1 Task 1: Early Detection of Signs of Depression

Table 3 summarizes the main steps of our proposed runs for the first task and the starting week for each run to make our first decisions.

Table 3: Summary of proposed runs for Task-1

	Step 1	Step 2	Step 3	Starting Week
LIRMMA	Doc2vec	Bayesian Inversion	MLP	8
LIRMMB	Word2vec	Bayesian Inversion	MLP	5
LIRMMC	Word2vec	Bayesian Inversion	RF	3
LIRMMD	Word2vec	Bayesian Inversion + Moving Average	———	1
LIRMME	Word2vec	Bayesian Inversion + Moving Average	———	1

For document vectorization (Doc2vec), the resultant vectors had 200 dimensions. The model used a context window of 10 words and a minimum of two for word counts in the text. It used a negative sampling loss with DBOW version and trained for 20 training epochs. In the word level vectorization, the vector size of a word had a dimension of 200 with context window size of five words. Hierarchical softmax was used and a minimum count of two words was considered.

In the second learning phase and for temporal modeling, the architecture of the Multi-layered perceptron (MLP) used had two hidden layers with ten neurons each. Concerning the Random Forest (RF) classifier, ten estimators were used.

LIRMMA started to give decisions from the eighth week. Positive users were reported for those with classification probability higher than 0.6. LIRMMB started to discover positive users from week five. The classification probability threshold varied from 1.0 in week five to 0.6 in the tenth week. LIRMMC started to give predictions from the third week. Since the small number of points in the time series from week three to five, we fixed the classification threshold to 1.0 and like LIRMMB, it started to reach 0.6 in the last week.

Hence we proposed LIRMMD and LIRMME to give a decision from the first week, we substitute the second learning phase with a window moving average from the output of the inversion technique. For LIRMMD, we assumed the positive users will have risky mood in the first chunks than the lasts. Two varying thresholds were used; one for the number of sentences and the other for the positive probability threshold. The size of averaging window is three and the probability changing from 0.6 with number of sentences higher than 100 to 0.8 and zero for sentences count threshold. For LIRMME, the difference comes from the assumptions that higher probability threshold was given to last weeks chunks than the first weeks with the same sentences thresholds. The risk probability starts with 0.8 in the first week to 0.6 in the last week.

5.2 Task 2: Early Detection of Signs of Anorexia

In this section, we highlight the difference between the proposed runs for anorexia second task. The steps are summarized in Table 4.

For text vectorization in document or word level, we used the same parameters as for the first task. The classification probability threshold for positive user was set to 0.8 in LIRMMA while in LIRMMB it was 0.6. For LIRMMC, LIRMMD and LIRMME, the same thresholds were used exactly as the same corresponding runs in the first task.

6 Results

6.1 Evaluation Results

Upon the submission of the last chunk, the evaluation process started for all runs results. As mentioned in Section 2, the tasks organizers use the two versions of

Table 4: Summary of proposed runs for Task-2

	Step 1	Step 2	Step 3	Starting Week
LIRMMA	Doc2vec	Bayesian Inversion	MLP	8
LIRMMB	**Doc2vec**	Bayesian Inversion	MLP	**8**
LIRMMC	Word2vec	Bayesian Inversion	RF	3
LIRMMD	Word2vec	Bayesian Inversion + Moving Average	———	1
LIRMME	Word2vec	Bayesian Inversion + Moving Average	———	1

ERDE in addition to the classical classification measures: Precision(P), Recall (R) and F1-Measure (F1). Tables 5 and 6 show the formal evaluation of all proposed runs for both tasks.

Table 5: Results of proposed runs for Task-1

	$ERDE_5$	$ERDE_{50}$	F1	P	R
LIRMMA	**10.66%**	9.16%	**0.49**	**0.38**	0.68
LIRMMB	11.81%	9.20%	0.36	0.24	**0.73**
LIRMMC	11.78%	9.02%	0.35	0.23	0.71
LIRMMD	11.32%	**8.08%**	0.32	0.22	0.57
LIRMME	10.71%	8.38%	0.37	0.29	0.52

6.2 Discussion

From the first look of the results, it is clear that document level vectorization behaves better than word level vectorization. But all runs using Doc2vec (LIRMMA for task-1 and LIRMMA & LIRMMB for task-2) started giving their decisions later than others. Also, runs that gave decisions from the first week give better ERDE with cut-off of 50 for both tasks. The runs with higher recall use word level vectorization with either MLP or RF as the second learning phase. In mode evaluation step, fake stories were misleading and made a lot of false positive predictions. In addition, our models does not discriminate between user posts and comments (posts without titles) which could be beneficial for evaluating user mood.

For some at-risk users, first chunks posts don't have any proof of depression or anorexia and suddenly users started to express their status late. For the second learning phase, the model classify the overall mood time series and late signs of disorders could not be predicted by our models. So, in some runs (for both

Table 6: Results of proposed runs for Task-2

	$ERDE_5$	$ERDE_{50}$	F1	P	R
LIRMMA	**13.65%**	13.04%	**0.54**	**0.52**	0.56
LIRMMB	14.45%	12.62%	0.52	0.41	0.71
LIRMMC	16.06%	15.02%	0.42	0.28	**0.78**
LIRMMD	17.14%	14.31%	0.34	0.22	0.76
LIRMME	14.89%	**12.69%**	0.41	0.32	0.59

tasks) some moderation on the proposed assumptions (classification probability thresholds) are needed.

Tables 7 and 8 show some statistics of all submitted runs compared to the proposed models. The ranking of the best run for each evaluation metric is also included. The statistics of the depression task are for 45 runs of 11 teams. The anorexia task statistics on results are for 34 runs of 9 teams. Most of the teams have participated in both tasks with at least one run for both. However the proposed architecture does not include any hand-crafted features or any attention signals, it seems to be comparable with other contributions for both tasks. The proposed models and runs act better for depression task than for anorexia task. The main reasons are that our models heavily depend on the data size and the leakage of anorexia data especially for positive users which was clear.

Table 7: Statistics on 45 participating runs results and our ranks

	$ERDE_5$	$ERDE_{50}$	F1	P	R
Average	10.33%	8.23%	0.42	0.37	0.55
Standard Deviation	1.13%	1.09%	0.12	0.15	0.16
Max	15.79%	11.95%	0.64	0.67	0.95
Min	8.78%	6.44%	0.18	0.1	0.15
Our Rank	31	22	13	15	3

7 Conclusion

In this paper we present the participation of LIRMM in the CLEF eRisk-2018 tasks. Both tasks are for early detection of signs of depression and anorexia from users posts on Reddit. We proposed five runs for each task and the results are, to some extent, interesting and comparable to other contributions. The proposed framework architecture used the text without any handcrafted features. It performs the classification through two phases of supervised learning using text

Table 8: Statistics on 34 participating runs results and our ranks

	$ERDE_5$	$ERDE_{50}$	F1	P	R
Average	13.31%	10.89%	0.56	0.63	0.58
Standard Deviation	1.62%	2.69%	0.19	0.22	0.2
Max	19.90%	19.27%	0.85	0.91	0.88
Min	11.40%	5.96%	0.17	0.15	0.1
Our Rank	28	27	20	24	4

vectorization methods. The first learning phase builds a time series representing the mood variation using the bayesian inversion technique. The second learning phase is a classification model that learns patterns from these time series to detect early signs of such mental disorders.

Predicting at-risk users once at the end for all runs was a good idea as some users expressed their disorders in the last chunks. But, some unreal stories in the firsts posts pushed the proposed runs to make a lot of false positive decisions. However the proposed architecture only uses the text blindly, adding some dictionary features or attention signals might have helped [17]. Another way is to build the second learner model much deeper using Convolution Neural Networks (CNN) [10] or Recurrent Neural Networks (RNN) architectures [17].

Acknowledgments

We would like to acknowledge La Région Occitanie and l'Agglomération Béziers Méditerranée which finance the thesis of Waleed Ragheb.

References

1. The national eating disorders association (NEDA).: Envisioning a world without eating disorders. In: The newsletter of the National Eating Disorders Association. Issue 22 (2009)
2. World Health Organization.: Depression and other common mental disorders: global health estimates. In: World Health Organization. http://www.who.int/iris/handle/10665/254610. License: CC BY-NC-SA 3.0 IGO (2017)
3. Bonsall, M.B., Wallace-Hadrill, S.M.A., Geddes, J.R., Goodwin, G.M., Holmes, E.A.: Nonlinear time-series approaches in characterizing mood stability and mood instability in bipolar disorder. In: Proceedings of the Royal Society of London B: Biological Sciences (2011)
4. Glenn, T., Monteith, S.: New measures of mental state and behavior based on data collected from sensors, smartphones, and the internet. In: Current Psychiatry Reports. vol. 16 (2014)
5. Hayda, A., Antoine, B., Marie-Jean, M.: Detecting early risk of depression from social media user-generated content. In: Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum. vol. CEUR-WS 1866 (2017)

6. Hoek, H.: Review of the worldwide epidemiology of eating disorders. In: Current Opinion in Psychiatry. vol. 29 (2016)
7. Jadayel, R., Medlej, K., Jadayel, J.J.: Mental disorders: A glamorous attraction on social media? In: Journal of Teaching and Education. vol. 07 (2017)
8. John, N., Kelly, A., Gregory., M., Jrgen, U., Lisa, M., Stephen, B.: Exploring opportunities to support mental health care using social media: A survey of social media users with mental illness. In: Early Intervention in Psychiatry (2017)
9. Joyce, D., L. Sulkowski, M.: The diagnostic and statistical manual of mental disorders: Fifth edition (dsm-5) model of impairment. In: Assessing Impairment: From Theory to Practice. pp. 167–189 (2016)
10. Kalchbrenner, N., Grefenstette, E., Blunsom, P.: A convolutional neural network for modelling sentences. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (2014)
11. Kishaloy, H., Lahari, P., Kan, M.Y.: Modeling temporal progression of emotional status in mental health forum: A recurrent neural net approach. In: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. pp. 127–135. Association for Computational Linguistics (2017)
12. Le, Q.V., Mikolov, T.: Distributed representations of sentences and documents. In: ICML. JMLR Workshop and Conference Proceedings, vol. 32, pp. 1188–1196. JMLR.org (2014)
13. Leite Barroso, M., Lucena Grangeiro Maranho, T., Melo teixeira batista, H., Pereira de Brito Neves, F., Farias de Oliveira, G.: Social panic disorder and its impacts. In: Amadeus International Multidisciplinary Journal. vol. 2, pp. 1–17 (2018)
14. Losada, D.E., Crestani, F.: A test collection for research on depression and language use. In: Conference Labs of the Evaluation Forum. pp. 28–39. Springer (2016)
15. Losada, D.E., Crestani, F., Parapar, J.: erisk 2017: Clef lab on early risk prediction on the internet: Experimental foundations. In: 8th International Conference of the CLEF Association. pp. 346–360. Springer Verlag (2017)
16. Losada, D.E., Crestani, F., Parapar, J.: Overview of eRisk – Early Risk Prediction on the Internet. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018). Avignon, France (2018)
17. Ma, Y., Peng, H., Cambria, E.: Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (2018)
18. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C.: Learning word vectors for sentiment analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1. pp. 142–150. HLT '11, Association for Computational Linguistics, Stroudsburg, PA, USA (2011)
19. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. In: Journal of Machine Learning Research. vol. 9, pp. 2579–2605 (2008)
20. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. In: CoRR. vol. abs/1301.3781 (2013)
21. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems 26. pp. 3111–3119. Curran Associates, Inc. (2013)

22. Mikolov, T., Yih, S.W.t., Zweig, G.: Linguistic regularities in continuous space word representations. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT-2013). Association for Computational Linguistics (2013)
23. Moulahi, B., Azé, J., Bringay, S.: Dare to care: A context-aware framework to track suicidal ideation on social media. In: Bouguettaya A. et al. (eds) Web Information Systems Engineering - WISE 2017.,Lecture Notes in Computer Science,. Springer, Cham. vol. 10570 (2017)
24. Patel, R., Lloyd, T., Jackson, R., Ball, M., Shetty, H., Broadbent, M., Geddes, J.R., Stewart, R., McGuire, P., Taylor, M.: Mood instability is a common feature of mental health disorders and is associated with poor clinical outcomes. In: BMJ Open. vol. 5. British Medical Journal Publishing Group (2015)
25. Paul, M.J., Dredze, M.: You are what you tweet: Analyzing twitter for public health. In: Adamic, L.A., Baeza-Yates, R.A., Counts, S. (eds.) ICWSM. The AAAI Press (2011)
26. Taddy, M.: Document classification by inversion of distributed language representations. In: CoRR. vol. abs/1504.07295 (2015)
27. Trautmann, S., Rehm, J., Wittchen, H.: The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? In: EMBO reports. vol. 17, pp. 1245–1249 (2016)
28. Trotzek, M., Koitka, S., Friedrich, C.: Linguistic metadata augmented classifiers at the clef 2017 task for early detection of depression. In: Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum. vol. CEUR-WS 1866 (2017)