Mining Twitter Data for Signs of Depression in Brazil

Otto von Sperling[1], Marcelo Ladeira[1]

Universidade de Brasília, Brazil
otto.vonsperling@aluno.unb.br, mladeira@unb.br

Abstract. The literature on computerized models that help detect, study and understand signs of mental health disorders from social media has been thriving since the mid-2000s for English speakers. In Brazil, this area of research shows promising results, in addition to a variety of niches that still need exploring. Thus, we construct a large corpus from 2941 users (1486 depressive, 1455 non-depressive), and induce machine learning models to identify signs of depression from our Twitter corpus. In order to achieve our goal, we extract features by measuring linguistic style, behavioral patterns, and affect from users' public tweets and metadata. Resulting models successfully distinguish between depressive and non-depressive classes with performance scores comparable to results in the literature. We hope that our findings can become stepping stones towards more methodologies being applied at the service of mental health.

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database Applications—*Data Mining*; I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems—*Medicine and Science*

Keywords: data mining, machine learning, mental health, twitter

1. INTRODUCTION

Mental illness has become a major cause of disability worldwide, with an estimate of 300 million people suffering from depression alone [World Health Organization 2017]. Although there have been growing efforts to treat and prevent mental illness through the use of technology, it is still a fairly recent field of research in Brazil.

The ubiquity of social media has provided researchers with a treasure trove of data, and the relationship that many have with their social profiles can lead to further insights into human behaviour by combining the power of data analysis with what is already known about mental health from rich bodies of research in psychology, psychiatry, neuroscience and sociolinguistics. It becomes possible to forecast the onset of depression [De Choudhury et al. 2013] and post-traumatic stress [Coppersmith et al. 2014] among other mental disorders.

Nevertheless, due to most research being carried out in English, often times there is a gap between the latest findings and how they translate to other cultures and languages. In light of such gap, we adapt well-established methods of data collection and feature extraction to build machine learning classifiers able to distinguish depressive from non-depressive users. Our results demonstrate that findings in the literature can be replicated and translated to the Brazilian culture. We aim to detect signs of Major Depressive Disorder [American Psychiatric Association 2000] in individuals on Twitter and how their collective behaviour differs from non-depressive individuals. Our main contributions with this article are:

(1) We construct a corpus with postings in Brazilian Portuguese from Twitter. The depressive class was built by extracting public messages of self-reported diagnosis from users (e.g., "I was diagnosed with depression") whilst the control group was built by using public self-reported messages of

positive mental states (e.g., "I am so happy today"). Additionally, we develop a data pipeline that enables us to both retrieve new users and update the current corpus.

(2) We extract features from user linguistic style and behavioral patterns to induce machine learning models that can distinguish between both depressive (positive) and non-depressive (control) classes.

(3) By showing that signals can be extracted from Twitter data in Brazil, and that data mining and machine learning can aid in detecting and studying mental disorders, we hope to inspire more future research in our country to approach the opportunity of developing technology to aid in the betterment of mental health of society at large.

For the sake of comprehension, we go over the structure of this article. Section 2 provides an overview of the milestones research in the intersection of mental health, sociolinguistics, data mining and machine learning has achieved over the last ten years. Section 3 describes the methods we employed; from data collection, to feature extraction, to machine learning models. Section 4 presents the results with information on the statistical significance of our features and the performance scores of our models, in addition to further considerations on our results. Section 5 makes the case for adapting methodologies developed for English corpora to other cultures and languages, Brazilian Portuguese in particular. Furthermore, we set the ground for future work to extend our corpus and improve our methods.

2. RELATED WORK

Since the early 2000s, there have been rising efforts to leverage the power of technology to aid in understanding and preventing mental disorders. From computerized analysis of written texts that revealed predictive cues about neurotic tendencies and psychiatric disorders [Rude et al. 2004], to support for the claim that negative (cognitive) processing biases in resolving ambiguous verbal information can predict subsequent depression [Williams and Galliher 2006], much has been accomplished in the past couple of decades. Nevertheless, the social media boom in the early 2010s brought with it an ever growing flux of data that enabled researchers to derive further insights from signals correlated to depressive and other mental disorders. Through the use of data mining techniques and time series analysis, research showed that patterns of behaviour can be matched to real world events [Bollen et al. 2011], and that symptomatic signals of major depressive disorder could be observed from status updates on Facebook [Moreno et al. 2011].

When it comes to Twitter, Park et al. 2012 found evidence that people post about their depression and treatment on the platform, and De Choudhury et al. 2013 induced Support Vector Machine (SVM) classifiers (precision = 0.742, recall = 0.629, F1 = 0.681) to estimate the risk of depression before its onset by measuring behavioral attributes relating to social engagement, emotion, linguistic styles, social network, and mentions of antidepressant medication. Coppersmith et al. 2014 proposed heuristics to automate parts of the corpus construction, which yielded, for depression alone, a corpus twice the size of De Choudhury et al. 2013, in addition to expanding the scope to other mental disorders. Following, Resnik et al. 2015 applied a variety of Supervised Topic Models on the corpus created by Coppersmith et al. 2014 and achieved promising results (AUC = 0.860) when classifying depressive versus non-depressive groups.

Following research questions the methods employed by De Choudhury et al. 2013 and argues that more meticulous methods are needed to support a stronger claim that Twitter data are capable not only of detecting depression, but can do so before the first diagnosis has been made [Reece et al. 2017]. Interpretability of models is yet another strong argument made by Reece et al. 2017, and Word Shift Graphs together with Hidden Markov Models (precision = 0.852, recall = 0.518, F1 = 0.644) are employed as an alternative that identifies signs not fully captured by either Linguistic Inquiry Word Count (LIWC) [Pennebaker et al. 2003] or Affective Norms for English Words (ANEW) [Bradley and

Lang 1999]. Having said all of that, claiming that the findings of Reece et al. 2017 deem invalid the methods of De Choudhury et al. 2013 is open for debate but it's rather unlikely. Conversely, it is to be seen as an iteration upon the methods to yield more robust claims of whether social media truly captures some of the nature of the human mind.

The one strong argument in this article is that instead of only making attempts to predict the chance of mental disorders in research corpus, comparable efforts should go into bridging the gap towards real-world tools that can aid practicing physicians, psychologists and care-takers better understand their patients and the patterns they share. Thus, with the present work, we hope to take our first steps towards the goal of detecting those who are in need now rather than later, and attract more research towards this challenge in our home country by showing that, although people and cultures differ in countless ways, leveraging technology in favor of well-being and mental health is truly a matter of grit and minds.

3. METHODS

In this research, we adapt the data collection method of Coppersmith et al. 2014 and the feature extraction method of De Choudhury et al. 2013. Our data collection pipeline is written in Python and uses Twitter developer's Application Programming Interface (API). This section spans from data collection, to feature extraction, to hyperparameter optimization for supervised learning algorithms.

3.1 Data Collection

All data collection took place between July, 2018 and May, 2019. Public messages posted between 2016 and 2019 have been collected for both depressive and non-depressive classes, for a total of 2941 Twitter users (N depre = 1486, N non-depre = 1455). In accordance to the Brazilian General Data Protection Regulation[1] and Twitter's Developer Policy[2], all of the data collected is public and has been anonymized for the sake of privacy. No information that can lead to participant identification will ever be made public regardless of whether its author's making it public themselves. No other person but the authors of this article have access to the data. No contact has ever been or will ever be made to participants, as the focus of this work is neither to diagnose nor to propose treatment to those in need at this time.

3.1.1 *Depressive Class (Positive).* We seek users who publicly state that they have been diagnosed with depression. We queried Twitter for public messages in Portuguese that contained self-reports of depression (e.g., "I was diagnosed with depression"). The reason why people come out publicly with such self-reports is presumably to seek support from the community, to explain some of their behaviour to their peers, or to fight against the stigma of mental illness. Some may jokingly or disingenuously make such statements, though the motivation behind such behaviour escapes the scope of this work. Nevertheless, we presume that the majority of self-reports is indeed truthful and will statistically overbear disingenuous ones. We then retrieve up to 3200 of the most recent public tweets for each user and routinely updated their message pool. Users with fewer than 30 messages in total or more than 300 messages in a single given day were filtered out and the remaining 1486 users were considered as positive examples. Both upper and lower bounds of 300 daily messages and 30 messages in total were put in place to remove supposedly spamming and marketing accounts and to have at least enough samples to enable some statistical analysis — although the latter follows a rule of thumb for statisticians rather than rigorous methods.

[1] Article 4(b) of Law 13.709/2018: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/Lei/L13709.htm
[2] Section F, item 2(b): https://developer.twitter.com/en/developer-terms/policy.html

3.1.2 *Non-Depressive Class (Control).* So as to be able to validate the data and to induce machine learning models that distinguish depressive from non-depressive signals, we queried Twitter for public messages in Portuguese that contained self-reports of positive mental states (e.g., "I am so happy with my new job!"). Further efforts were made to remove allegedly depressive users from the non-depressive group by filtering out any users that report diagnosis of depression and also anxiety, due to observed comorbidity between both disorders [Sartorius et al. 1996]. Next, we retrieve up to 3200 of the most recent public tweets for each user, routinely updated their message pool, and filter out users with fewer than 30 messages in total or more than 300 messages in a single given day. The remaining 1455 users were considered as negative examples.

3.1.3 *Caveats.* No effort was applied in verifying either the veracity or the onset date of users' depressive disorder from self-report of diagnosis. It is often the case that no mention is made of when the diagnosis was first reached or whether it was the first diagnosis at all. Some users even go as far as reporting diagnosis from childhood and adolescence. For such cases, we conjecture that the high rate of recurrence of depression, which reflects an underlying vulnerability that is both largely genetic in nature as well as due to psycho-social risk factors [Burcusa and Iacono 2007], can be used to interpret self-reports of diagnosis as an arguably reliable statement of relapse. Thus, behavioral and linguistic style attributes of such users is assumed to more closely match those of users in the depressive class. It also follows that we establish no time-wise limit to when messages were posted other than the span between 2016 and 2019.

3.2 Feature Extraction

We extract 5 categories of attributes, for a total of 8 attributes, from the Twitter posts collected in the span between 2016 and 2019. Our attributes are defined as follows:

—Engagement: we define (1) *volume* as an attribute measured by the total number of messages per user, per day.
—Linguistic Style: simple natural language processing techniques were used to derive three attributes regarding pronoun use. (2) *fpp* is defined as the daily count of first-person pronouns in posts. The next two attributes follow the same logic for (3) *spp* second-person pronouns and (4) *tpp* third-person pronouns.
—Emotion: the unigram sentiment instrument ANEW-Br [Kristensen et al. 2011] was used to derive 2 attributes in this category. (5) *valence*, which ranges from unpleasant to pleasant, and (6) *activation*, which ranges from relaxed to tense (e.g., sadness and anger, with low and high activation respectively and low valence for both).
—Depression Terms: We define (7) *depre_terms* as the ratio of the number of words in a message that belong to the depression lexicon for Brazilian Portuguese [Nascimento et al. 2018] to the total number of words in a message.
—Insomnia Index: Insomnia has been shown to have significant correlation with depression [Jansson-Fröjmark and Lindblom 2008]. Thus, we define (8) *insomnia index* as the daily ratio of messages posted at night ("11PM-6AM") to messages posted during the day ("6AM-11PM").

Feature vectors were constructed for each user, for a total of 32 features. Each of the 8 attributes was used to generate one time series, which in turn yielded 4 features defined as follows:

(1) *Mean frequency* (μ_i) as the average measure of the time series signal of an attribute over the entire period of analysis
(2) *Variance* as the variation in the time series signal over the entire time period. Given a time series $X_i(1), X_i(2), ..., X_i(t), ..., X_i(N)$ on the i^{th} measure, it is given as:

$$(1/N)\Sigma_t(X_i(t) - \mu_i)^2 \quad (1)$$

(3) *Mean momentum* as the relative trend of a time series signal, compared to a fixed period before. Given the above time series, and a period length of M (=7) days, its mean momentum is:

$$(1/N)\Sigma_t(X_i(t) - (1/(t-M))\Sigma_{(M \leq k \leq t-1)}X_i(k)) \quad (2)$$

(4) *Entropy* as the measure of uncertainty in a time series signal. For the above time series it is:

$$-\Sigma_t X_i(t) log(X_i(t)) \quad (3)$$

Thus, each user is represented by a standardized 32-item feature vector with zero *mean* and unit *variance*.

3.3 Machine Learning Models

We train supervised machine learning classifiers to discriminate between depressive and non-depressive classes. The corpus is randomly split into training (50%), validation (20%) and test (30%) sets, the latter being used to report the average performance scores over of our the models. Based on the principle of Ockham's razor, which states that the simplest solution is most likely the right one, and to avoid overfitting, we employ principal component analysis (PCA) to reduce dimensionality and capture 96% of the variance in our dataset, which in turn yields 13 principal components (features) — namely, *fpp_var*, *volume_mean*, *volume_entropy*, *insomnia_var*, *insomnia_momentum*, *valence_mean*, *valence_var*, *valence_momentum*, *valence_entropy*, *activation_mean*, *activation_var*, *activation_momentum*, *activation_entropy*. Furthermore, we use the *scikit-learn*[3] Python module to construct three distinct classifiers.

—*Support Vector Machine* (SVM): We employ grid search over 5,625 hyperparameter combinations with 10-fold cross-validation to optimize our Support Vector Machine classifier. Our best performing SVM classifier has the hyperparameters *gamma* set to 'scale' and C to 45.

—*Random Forest* (RF): Next, we utilize 100 iterations of randomized search with 5-fold cross-validation to optimize our Random Forests hyperparameters. Our best set of hyperparameters is a 250-tree Random Forests classifier with maximum depth of 15 and a minimum number of 20 samples required to split an internal node.

—Multilayer Perceptron (MLP): Finally, we take advantage once again of grid search to find optimal settings over 2,500 hyperparameter combinations, which yields a thirteen-node two-hidden-layer MLP classifier with the hyperparameters *alpha* and *momentum* set to $1e-2$ and $9e-1$ respectively.

All models are compared to a stratified dummy classifier that generates predictions by respecting the training set's class distribution. Given that both classes are closely matched, the dummy classifier behaves as random chance over a large number of iterations.

4. RESULTS

We first present results of the statistical significance of our features in 5 of the 8 attribute classes extracted. We use independent sample t-tests to compare the mean of the depressive and non-depressive classes for p-value $\leq \alpha = 0.05/32 = 1.5625e-3$, after Bonferroni correction for multiple comparisons. The values of the t-statistics, which represent the sensitivity of each feature, and the corresponding p-values are presented in Table I.

Our statistical results align with De Choudhury et al. 2013 to some extend. Variance is the attribute that sees the highest number of statistically significant features, although we argue that mean and momentum are still more relevant. The higher overall t-values for features derived from both

[3] https://scikit-learn.org/stable/index.html

Table I. Statistical significance of principal components

Attributes	**Entropy**	**Mean**	**Momentum**	**Variance**
activation	4.6257 **	10.3910 **	9.3983 **	6.367147 **
fpp	0.1476	0.9533	1.3704	4.014316 **
insomnia	0.4773	0.2027	12.1403 **	4.018704 **
valence	4.6273 **	8.7854 **	7.9036 **	4.545939 **
volume	3.9258 **	5.7723 **	3.0485	1.050110

$df = 2939$ ** $p \leq \alpha$, after Bonferroni correction

measures shows that they bear a heavy weight in rejecting the null hypothesis that the two classes — depressive and non-depressive — have similar signals of behaviour, language and cognition. It is a fact that we extract fewer signals in comparison, although we argue that major underlying signals of depression can be captured from signs of insomnia [Jansson-Fröjmark and Lindblom 2008], high attention to self [Rude et al. 2004], lack of energy and exhaustion [Rabkin and Struening 1976], loss of social connectedness [Williams and Galliher 2006], and elevated negative emotion [Park et al. 2012]. Thus, our choice of features seems adequate and yields positive results in classification.

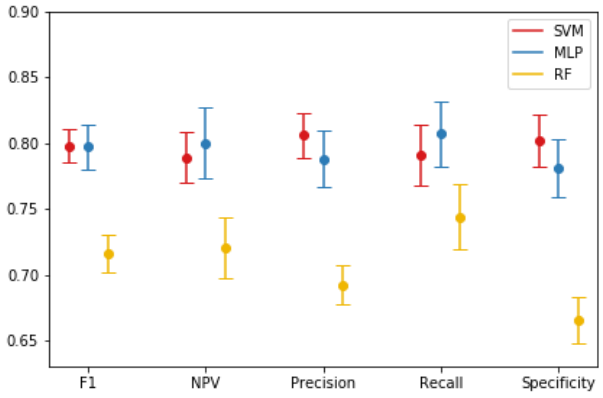
Machine Learning Classifiers We choose to go beyond the usual set of performance scores for the sake of better understanding and comparison of our models. The average precision, recall, specificity, negative predictive value (NPV), and F1 scores are reported in Table II with 95% confidence intervals (CI) over 100 randomized runs. Our best classifiers were the Multi-layer Perceptron and the Support Vector Machine due to their higher chance of detecting most users who indeed suffer from depression (recall), in addition to a higher negative predict value (i.e., there is a higher chance that all non-depressive users will be detected). Overall, both the SVM and MLP classifiers have comparable performance, as shown by the intersecting confidence intervals of their classification performance scores in figure 1. We opt to prioritize recall over other measures since such computerized tools are not meant to supplant professional diagnosis. Instead, they should enable physicians and care-takers to spend more time in direct contact with individuals who do indeed need their care.

Table II. Average performance scores over 100 randomized runs with 95% confidence interval

	F1	npv	precision	recall	specificity
SVM	0.798 ± 0.0123	0.789 ± 0.0193	0.806 ± 0.0170	0.791 ± 0.0227	0.802 ± 0.0201
MLP	0.797 ± 0.0175	0.800 ± 0.0269	0.788 ± 0.0212	0.807 ± 0.0251	0.781 ± 0.0223
RF	0.716 ± 0.0144	0.720 ± 0.0230	0.692 ± 0.0147	0.744 ± 0.0248	0.665 ± 0.0175
BASELINE	0.502 ± 0.0180	0.492 ± 0.0214	0.502 ± 0.0180	0.502 ± 0.0180	0.493 ± 0.0129

Both the SVM and MLP classifiers improved over De Choudhury et al. 2013, though the comparison is loosely formal and some considerations have to be made. First, De Choudhury extracts features from 1 year of user history seeking to predict the onset of MDD, whereas we do so in a span of 3 years to detect users with overall observed behaviour correlated with depression in the literature; thus, direct comparisons of complementary and yet distinct objectives is unsuited under rigorous analysis. Conversely, no effort has been made in this work to verify whether users in fact suffered from MDD. De Choudhury employs both CES-D (Center for Epidemiologic Studies Depression Scale) [Radloff 1977] and Beck Depression Inventory (BDI) [Beck et al. 1996] to screen participants for MDD. Nonetheless, Coppersmith et al. 2014 demonstrates that trading off meticulous screening for larger amounts of data not only yields comparable results, but also enables more robust machine learning techniques to be leveraged. Hence, we present our results with confidence despite the aforementioned caveats.

Fig. 1. Comparative plot of interval confidence for the classifiers' performance scores



To some extend, our challenge differs in nature from Coppersmith et al. 2014 and De Choudhury et al. 2013. The cultural and language barrier takes its toll on natural language processing techniques as well as on the availability and accuracy of lexicons. For instance, the depressive lexicon for Brazilian Portuguese [Nascimento et al. 2018] failed to show statistical significance at this time, whereas Coppersmith et al. 2014 demonstrates the efficacy of using lexicons to construct novel methodologies. On the other hand, all four features derived from the ANEW-Br lexicon [Kristensen et al. 2011] reached statistical significance, which in turn shows that, although nascent, this field of research is likely to thrive in Brazil.

5. CONCLUSION AND FUTURE WORK

The aim of the present work was to demonstrate the feasibility of adapting methodologies designed to tackle data collection, feature extraction and induction of machine learning models from English corpora, to apply them in Brazil and identify signs of depression based on users' Twitter data. Our findings strongly support the claim that computational methods can effectively screen Twitter data for indicators of depression in a Portuguese corpus with performance comparable to previous findings for English corpora. We **1)** constructed a robust corpus through an automated data pipeline built with Python, **2)** extracted features based on the literature of psychology, psychiatry and sociolinguistics that capture some of the underlying signals of both depressive behaviour and language with statistical significance, and **3)** induced classifiers to distinguish the depressive from non-depressive class, in addition to presenting the performance scores (with 95% CIs) that compare to, if not improve upon, the results in the literature [Park et al. 2011; De Choudhury et al. 2013; Coppersmith et al. 2014]. By doing so, we demonstrate the usefulness of social media to detect relevant traces of behaviour and users' state of mind in Brazil, and showed that despite cultural differences, there is an underlying pattern to the hidden signs people leave behind.

As future work, we propose **1)** extending the feature set extracted from Twitter data — and other social media — to better capture classes of signals that are unique to Brazilians, **2)** adapting our methods to apply across several nationalities; researchers could benefit from the ability to draw comparative studies, and thus detect what signals are shared by all nationalities, and the signals that are unique to a smaller population. In fact, city-, state-, and nation-wide lenses can be employed to better understand local trends and draw more effective actions by governments, physicians, psychologist, caretakers and the public in general. Furthermore, we suggest **3)** applying unsupervised learning algorithms, such as long short-term memory (LSTM) recurrent neural networks, to corpora in Portuguese, in addition to **4)** word shift graphs and other methods that yield further interpretability of models' outcomes.

In the future, we hope to see much more research in mental health and data mining in our home country. Regrettably, far too often technology is taken for granted or disingenuously employed. One of the major driving forces of innovation in natural language processing, behaviour modeling, trend prediction, and others alike — *profit maximization* — is not usually aligned with the well-being of the public at large. Despite the diversion, public interest has been shifting towards a more open (and equally difficult) discourse about privacy and individual rights. In the scope of individual rights, we argue that mental health, emotional stability and supportive care are intrinsic to having a fulfilling journey through life; thus, an individual right.

REFERENCES

American Psychiatric Association. pp. 0–942. In , *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition: DSM-IV-TR*. American Psychiatric Association, Michigan, USA, pp. 0–942, 2000.

Beck, A. T., Beck, R. A., and Brown, G. K. Manual for the beck depression inventory-ii. *Psychological Corporation* 78 (2): 490–498, 1996.

Bollen, J., Mao, H., and Pepe, A. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Fifth International AAAI Conference on Weblogs and Social Media*. Vol. 2011. AAAI Conference, Palo Alto, USA, pp. 450–453, 2011.

Bradley, M. M. and Lang, P. J. Affective norms for english words (ANEW): Instruction manual and affective ratings. *The Center for Research in psychophysiology* 30 (1): 25–36, 1999.

Burcusa, S. L. and Iacono, W. G. Risk for recurrence in depression. *Clinical psychology review* 27 (8): 959–985, 2007.

Coppersmith, G., Dredze, M., and Harman, C. Quantifying mental health signals in twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*. Association for Computational Linguistics, Baltimore, USA, pp. 51–60, 2014.

De Choudhury, M., Gamon, M., Counts, S., and Horvitz, E. Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*. AAAI conference on weblogs and social media, Cambridge, USA, pp. 0–10, 2013.

Jansson-Fröjmark, M. and Lindblom, K. A bidirectional relationship between anxiety and depression, and insomnia? a prospective study in the general population. *Journal of Psychosomatic Research* 64 (4): 443 – 449, 2008.

Kristensen, C. H., de Azevedo Gomes, C. F., Justo, A. R., and Vieira, K. Brazilian norms for the affective norms for english words. *Trends in Psychiatry and Psychotherapy* 33 (3): 135–146, 2011.

Moreno, M. A., Jelenchick, L. A., Egan, K. G., Cox, E., Young, H., Gannon, K. E., and Becker, T. Feeling bad on facebook: Depression disclosures by college students on a social networking site. *Depression and anxiety* 28 (6): 447–455, 2011.

Nascimento, R. S., Parreira, P., Santos, G. N., and Guedes, G. P. Identifying signs of depressive behaviour on social media (identificando sinais de comportamento depressivo em redes sociais). In *7º Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2018)*. SBC, Porto Alegre, Brazil, pp. 0–6, 2018.

Park, M., Cha, C., and Cha, M. Depressive moods of users portrayed in twitter. In *Proceedings of the ACM SIGKDD Workshop on healthcare informatics (HI-KDD)*. Vol. 2012. ACM SIGKDD, Philadelphia, USA, pp. 1–8, 2012.

Pennebaker, J. W., Mehl, M. R., and Niederhoffer, K. G. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology* 54 (1): 547–577, 2003.

Rabkin, J. G. and Struening, E. L. Life events, stress, and illness. *Science* 194 (4269): 1013–1020, 1976.

Radloff, L. S. The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement* 1 (3): 385–401, 1977.

Reece, A. G., Reagan, A. J., Lix, K. L., Dodds, P. S., Danforth, C. M., and Langer, E. J. Forecasting the onset and course of mental illness with twitter data. *Scientific reports* 7 (1): 13006, 2017.

Rude, S., Gortner, E.-M., and Pennebaker, J. W. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion* 18 (8): 1121–1133, 2004.

Sartorius, N., Üstün, T. B., Lecrubier, Y., and Wittchen, H.-U. Depression comorbid with anxiety: results from the WHO study on psychological disorders in primary health care. *The British journal of psychiatry* 168 (S30): 38–43, 1996.

Williams, K. L. and Galliher, R. Predicting depression and self-esteem from social connectedness, support, and competence. *Journal of Social and Clinical Psychology - J SOC CLIN PSYCHOL* 25 (8): 855–874, 10, 2006.

World Health Organization. Depression and other common mental disorders: Global health estimates. https://bit.ly/30iFz52, 2017.