

Original Paper

Validating Machine Learning Algorithms for Twitter Data Against Established Measures of Suicidality

Scott R Braithwaite^{1*}, PhD; Christophe Giraud-Carrier^{2*}, PhD; Josh West^{3*}, PhD; Michael D Barnes^{3*}, PhD; Carl Lee Hanson^{3*}, PhD

¹Computational Health Science Research Group, Department of Psychology, Brigham Young University, Provo, UT, United States

²Computational Health Science Research Group, Department of Computer Science, Brigham Young University, Provo, UT, United States

³Computational Health Science Research Group, Department of Health Science, Brigham Young University, Provo, UT, United States

* all authors contributed equally

Corresponding Author:

Carl Lee Hanson, PhD

Computational Health Science Research Group

Department of Health Science

Brigham Young University

4103B Life Sciences Building

Provo, UT, 84602

United States

Phone: 1 (801) 422 9103

Fax: 1 (801) 422 0004

Email: carl_hanson@byu.edu

Abstract

Background: One of the leading causes of death in the United States (US) is suicide and new methods of assessment are needed to track its risk in real time.

Objective: Our objective is to validate the use of machine learning algorithms for Twitter data against empirically validated measures of suicidality in the US population.

Methods: Using a machine learning algorithm, the Twitter feeds of 135 Mechanical Turk (MTurk) participants were compared with validated, self-report measures of suicide risk.

Results: Our findings show that people who are at high suicidal risk can be easily differentiated from those who are not by machine learning algorithms, which accurately identify the clinically significant suicidal rate in 92% of cases (sensitivity: 53%, specificity: 97%, positive predictive value: 75%, negative predictive value: 93%).

Conclusions: Machine learning algorithms are efficient in differentiating people who are at a suicidal risk from those who are not. Evidence for suicidality can be measured in nonclinical populations using social media data.

(*JMIR Mental Health* 2016;3(2):e21) doi: [10.2196/mental.4822](https://doi.org/10.2196/mental.4822)

KEYWORDS

suicide; social media; twitter; machine learning

Introduction

Suicide claims more than twice as many lives each year as compared with homicide, and is the 10th leading cause of death in the United States (US). It is found that the suicide count rises above 33,000 every year [1] and more than 30 people attempt suicide following each death [1]. This eventually results in emotional and financial burdens on their families and loved ones. The World Health Organization [2] recently endorsed

several universal interventions to prevent suicide, of which two promising strategies were to target vulnerable groups and individuals and then facilitate their access to crisis helplines. Timely identification of these vulnerable groups and individuals [3], and the balance of identifying high-risk cases without too many false positives [4,5], however, remains a challenge. This has led to increasing efforts in clinical settings, which further increase financial and time-related costs [6]. Due to these reasons, the public health priority is to explore novel approaches and identify individuals at risk for suicide without increasing

costs or adding burdens to the already present clinical system. This effort may benefit from the introduction and proliferation of emerging social media technologies.

Social media has provided researchers with new avenues to employ automated methods for analyzing language and to better understand individuals' thoughts, feelings, beliefs, behavior, and personalities [7]. Studies of language-using computational data-driven methodologies have demonstrated utility for monitoring psychological states and public health problems such as influenza [8,9], heart disease mortality [10], drinking problem [11], prescription drug abuse [12,13], and tobacco use [14]. Infodemiology or infoveillance is an emerging field related to computational data-driven methodologies and other studies, that use social media to understand and monitor health problems efficiently [15,16].

Twitter is a social media application that allows users to broadcast news, information, and personal updates to other users (followers) in tweets or statements of 140 characters or less. Speech is considered to be an important marker for both depression and suicide risk assessment [17], and Twitter provides a novel social media avenue for exploring these major public health concerns. Initial studies confirmed the role of infodemiology or infoveillance for social media data to track health trends [8-14], something beyond its original purpose. However, these early studies are limited as they failed to prove that the observed trends reflect the actual values. The advanced research in several studies has focused on confirming social media observations, which increases our collective confidence in these data values that act as a source for monitoring health issues and trends [18-20]. Most relevant to the work presented here, recent studies have used Twitter, Sina Weibo (Sina Weibo is a Chinese social media and microblogging site similar to Twitter), and Reddit specifically to tackle suicidality [21-25]. An annual workshop series commenced in 2014 on *Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, has attracted a host of social media data-driven work on suicide, as well as other mental health issues, including depression, schizophrenia, dementia, and posttraumatic syndrome disorder (eg, see [26-30]). Additional research remains warranted to demonstrate the safety and efficacy of social media prevention activities [31], and methodological issues need further refinement especially in terms of specificity and sensitivity of suicide risk. The purpose of this study was to validate the use of machine learning for Twitter data against empirically validated measures of suicidality in the US population with an eye for suicide prevention.

Methods

Recruitment and Procedure

The participants for this study were selected through Amazon's Mechanical Turk (MTurk, www.mturk.com). Participants in the US who were frequent Twitter users and above 18 years of age were invited to participate in a "Survey for Twitter users (~10 min)." Only those who had completed more than 500 Human Intelligence Tasks (HITs)—the name MTurk gives to online tasks, including surveys, transcriptions, categorization of receipt items, etc.—with an approval rate of > 95% (ie,

requesters found their work was acceptable for more than 95% of tasks they had undertaken) were allowed to complete the survey. Participants were informed that this survey was for Twitter users and that "Only those who are active Twitter users with public, personal Twitter accounts may participate, we will not approve any workers who do not meet these qualifications." To ensure the eligibility of participants, they had to complete a screening questionnaire before accepting the HIT. The screening questionnaire questioned whether they had an active, public Twitter account, how long the account had been active, and how often they tweeted. Our survey was published during the early summer of 2014 and republished during the early fall of 2015. Participants were paid according to the current MTurk market rates (ie, between 30 and 50 cents). The authors' university institutional review board approved all the study procedures and measures.

Stimuli: Human Intelligence Tasks

Participation in the study consisted of providing a Twitter handle and completing a set of questionnaires that assessed psychosocial functioning. A Twitter handle is a username that takes the form @username. The questionnaires examined in the present study are the Depressive Symptom Inventory–Suicide Subscale (DSI-SS), The Interpersonal Needs Questionnaire (INQ), and Acquired Capability for Suicide Scale (ACSS). The DSI-SS, a 4-item screening tool for suicidal symptoms assesses suicidality in a reliable and valid manner. In addition to an established clinical cutoff, it assesses for resolved plans and presuicidal actions, which are absent in most suicidal cases [32]. The INQ and ACSS scales assess facets of Joiner's Interpersonal Theory of suicide: thwarted belongingness, perceived burdensomeness (INQ), and the acquired capability for suicide (ACSS) [33]. These scales have demonstrated good reliability and construct validity [34].

Participants

In Summer 2014, we decided to obtain a sample of 100 participants. Beginning with 489 potential participants, we dropped 251 participants that did not actually provide data (most of these were likely bots, which are computer programs designed to generate responses to HITs in hopes of receiving payment). Researchers studying MTurk data collection recommended involving high reputation participants (those with a high number of completed and approved HITs) and including attention control checks [35]. We included control questions to ensure that those who responded were providing reliable data. Our control questions were designed to discern whether the participant was paying attention to each question (eg, "In the last month how often have you showed that you were paying attention by selecting 'Sometimes'"). We included five control questions; participants who failed two or more were excluded. About 46 participants who failed to answer the control questions were excluded. Finally, five participants attempted the survey more than once, in some cases with variable answers. Since, it was impossible to decide which of their answers was valid, we removed these respondents and their duplicates (17 participants in total), resulting in 175 participants. To validate the Twitter handles provided by the corresponding MTurk participants, we used the Twitter API via Twitter4J, and queried the last status

by the handle. We removed all users who could not answer the query as it indicates that either the user does not exist or the user's account is not public. As individuals sometimes find it funny to mention a celebrity handle as their own, we verified that the user was not a celebrity. If an account was verified as a celebrity account, we removed the corresponding user. In addition, to ensure that the MTurk participants had some activity on Twitter and that they would have a sufficient number of tweets for our analysis, we removed all users who posted less than two posts per month on an average. To find the average number of posts per month, we divided the total number of

tweets that the user posted by the number of months, since the user's account was created. Finally, we removed all users whose last tweet was more than 1 month old. In total, there were 101 MTurk workers with both permissible responses and exploitable Twitter accounts. In Summer and Fall 2015, we repeated the above data collection procedure to extend the size of our sample. We began with 111 more potential candidates, excluded 77 of them, resulting in an additional 34 exploitable Twitter accounts for a final sample of size $N=135$. Participant characteristics are shown in Table 1. For all valid user accounts, we again queried the Twitter API to collect the latest 200 tweets of the user.

Table 1. Participant characteristics

Ethnicity	
African American	19 (14.1%)
Asian	5 (3.7%)
Latino	6 (4.4%)
Mixed/Biracial	9 (6.7%)
Caucasian (White)	95 (70.4%)
Native American	1 (0.7%)
Education	
Graduate of professional degree	14 (10.4%)
Bachelor's degree	48 (35.6%)
Some college	60 (44.4%)
High school or equivalent (eg, GED)	9 (6.7%)
Less than high school	4 (2.9%)
Income	
Over \$150K	2 (1.5%)
\$100K–\$150K	4 (3.0%)
\$75K–\$100K	12 (8.9%)
\$50K–\$75K	28 (20.7%)
\$25K–\$50K	41 (30.3%)
Under \$25K	46 (34.1%)
None	2 (1.5%)
Twitter account creation date	
2008	14 (10.4%)
2009	39 (28.9%)
2010	17 (12.6%)
2011	23 (17.1%)
2012	20 (14.8%)
2013	13 (9.6%)
2014	8 (5.9%)
2015	1 (0.7%)

^aGED General Educational Development

Analysis of Tweets

For each participant, the textual content of all of the retrieved tweets was aggregated into a single file. Each file was then analyzed with the updated 2015 version of Linguistic Inquiry and Word Count software (LIWC) [36]. LIWC is a language analysis tool that extracts information from text in three main forms. The first, new in the 2015 version, consists of four variables that capture global high-level properties of the text as percentiles, namely, analytical thinking, clout, authenticity, and emotional tone. The second consists of 71 variables that represent the relative use of predefined word categories, from linguistic forms, such as pronouns and verbs, to psychological, social, emotional, and cognitive mechanisms, such as family, anger, sadness, certainty, leisure, religion, and death. The third focuses on the relative use of 17 language markers (eg, swear words, netspeak) and punctuation categories. For each of its 88 base categories, LIWC computes the percentage of total words in that category within the body of text being analyzed. For example, if a text sample has 125 words, and 3 of these words belong the pronoun category, LIWC gives a score of 2.4 (3/125) to that category. LIWC has been validated in a number of studies in the context of social media data [37,38] and has been previously proved to correlate in meaningful ways with suicidality [39-41]. LIWC has also been used to annotate tweets as showing signs of distress, where distress is regarded as a risk factor for suicide [42], as well as to analyze language differences across ten mental health issues [43].

Description of Sample

Our sample consists of 85 females and 50 males with an ethnic composition largely consistent with the US population, with a slight under-representation of Latino, and overrepresentation of mixed/biracial individuals. The distribution level of education and income suggests that our sample consists of generally more educated and affluent individuals than the national average, consistent with the findings of other researchers [44]. About 7 (5.2%) individuals were identified as homosexual, 18 (13.3%) as bisexual, and 110 (81.5%) as heterosexual. The proportion of individuals who were not heterosexual is higher than that expected from population norms, perhaps because social media provides readier access to lesbian, gay, bisexual, and transgender (LGBT) populations than traditional methods of sampling [45]. All Twitter accounts were listed as English accounts, and most users had been active for several years as shown by the distribution of account creation dates in Table 1. Almost half of the users (64 of 135) had posted over 2000 tweets at the time of data collection. About 17 individuals in our sample could be confidently considered as clinically significantly suicidal, since their DSI-SS score was greater than 2; the remaining 118 individuals were deemed nonsuicidal.

Machine Learning

For each participant, we built a feature vector consisting of the LIWC variables, together with a target class label: suicidal, nonsuicidal (as determined by the DSI-SS). The set of 135 vectors form a training data set that can be used by classification learning algorithms to induce a predictive model of suicidality. We implemented the predictive analysis in Python, using the scikit-learn library [46].

Various classification learning algorithms are available. For this study, our aim was to build not only a model with good accuracy but also one that would potentially provide insight into its predictions. Hence, decision tree learning was selected as it empirically produced accurate models for a large number of applications, and the models it built are easily interpretable [47]. Decision tree learning implements a kind of divide-and-conquer approach to learning. At each stage, a feature is selected and becomes the root of a subtree whose branches are the values, or ranges of values, of the selected feature. The training data are partitioned along these values, and sent down the corresponding branch. The process is then applied recursively to each partition until all training examples in the partition have the same label or there are no more features to select; at this point, a leaf node is created and labeled with the most prevalent label in the partition. A new example is classified by starting at the root of the tree, and following a path to a leaf node such that at each internal node the example takes the branch corresponding to its value for the feature at that node. The leaf node label is the predicted label for the new example. Note that the prevalence computed for a leaf node during training may in turn serve as a measure of confidence in its predictions. During learning, feature selection is affected by maximizing gain of information or the difference between the entropy of the training data before and after partitioning. Entropy measures the purity of a set of training examples with respect to the class label. A set where all of the examples have the same label has minimum entropy, while a set where the examples are spread uniformly over all labels has maximum entropy. Hence, at each stage, the attribute that is best at discriminating among the training examples at that stage is selected.

We estimate the accuracy of our decision tree learning approach using leave-one-out cross-validation (loo-cv), wherein, a decision tree is induced from all but one of the participants' feature vectors and tested on the out-of-training participant. The process is repeated N times, until each participant has been left out for testing. For each participant, we record whether the prediction was correct and aggregate over all participants to obtain an overall accuracy value.

Results

Analysis of responses to the INQ across the entire group and for each subgroup (suicidal and nonsuicidal, as defined by the DSI-SS cut-point) revealed significant differences. Suicidal individuals endorsed significantly less belongingness (one-tailed independent sample $t(133) = -5.84, p < .001$; Cohen's $d = -1.52, 95\% CI [-2.05, -0.97]$), and significantly higher burdensomeness ($t(133) = -8.41, p < .001$; $d = -2.18, 95\% CI [-2.75, -1.61]$). Those indicating significant suicidality reported a slightly higher acquired capability for suicide ($t(133) = -1.91, p = .03$; $d = -0.49, 95\% CI [-1.00, 0.19]$). These results offer additional support for the validity of the INQ and provide converging evidence of the suicidality of those who were above the cutoff on the DSI-SS.

If the DSI-SS cutoff identifies 17 individuals in our sample as suicidal, then the default accuracy of a predictive model, obtained by indiscriminately predicting the most prevalent class

(here, nonsuicidal), is 87.4% (118/135). The decision tree's loo-cv accuracy was 91.9%. The confusion matrix, shown in

Table 2, gives rise to the following values.

Table 2. Loo-cv confusion matrix for decision tree learning

	Suicidal	Not suicidal
Suicidal	9	8
Not suicidal	3	115

- Sensitivity: 0.53 (8 suicidal individuals were wrongly labeled as nonsuicidal)
- Specificity: 0.97 (only 3 out of 118 nonsuicidal individuals were wrongly labeled as suicidal)
- Positive predictive value: 0.75 (only 3 of the 12 individuals labeled as suicidal were actually not suicidal)
- Negative predictive value: 0.93 (only 8 of the 123 individuals labeled as nonsuicidal were actually suicidal)

The pruned decision tree induced from the complete sample is shown in Figure 1. It is included here for its explanatory power.

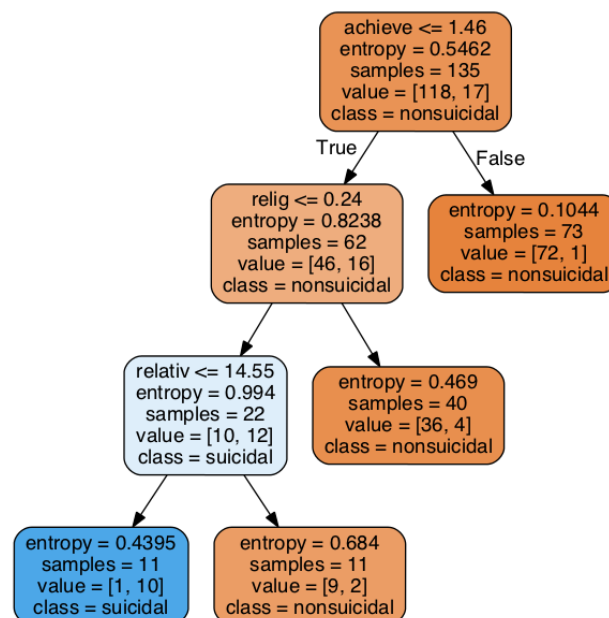
We note that there were minor differences in deeper parts of the unpruned trees induced over various runs, as increased depth tends to lead to overfitting. However, the macrostructure of the pruned tree depicted in Figure 1 remains consistent across runs, suggesting that the results should generalize.

The structure of the tree is rather consistent with intuition as well. The tree first splits on the “achieve” category of LIWC, such that if an individual's usage rate of achievement-related words exceeds 1.46, that individual is labeled as nonsuicidal. It is striking that the corresponding leaf node has very low entropy, indicating that 72 of the 73 individuals in our sample

satisfying the condition are indeed nonsuicidal. A noted fact is that a value of 1.46 for the “achieve” category is larger than the mean “achieve” values of most genres of writing analyzed using LIWC, as reported in the LIWC documentation [36]. This suggests that relative to others, these individuals' tweets have a higher proportion of achievement-related words, and that this high degree of achievement talk covaries with nil levels of clinically significant suicidality.

The next node where the tree splits (left branch) contains the “religion” category of LIWC. If an individual's rate of usage of religion-related words exceeds 0.24, then that individual is labeled as nonsuicidal. As seen above, it is striking that the corresponding leaf node has rather small entropy, giving relatively high confidence to the prediction (90%; 36 of 40). This seems to confirm other studies suggesting that religiosity may act as a protective factor against depression, social isolation, and suicidality. If the rate of religion-related words is low, the prediction of suicidality jumps to just about 50% (12 of 22). The final split of the pruned tree contains the “relativity” category, which is related to notions of motion, space, and time. It provides a rather clean separation between suicidal and nonsuicidal individuals.

Figure 1. Result from Decision Tree Learning Algorithm.



Discussion

Principal Findings

Suicide continues to be one of the leading causes of death in the US [1] and new methods of assessment capable of tracking

suicide risk in real time are required. Our findings reveal that machine learning algorithms can help differentiate between those who are at suicide suicidal risk and those who are not. Below we discuss these findings in light of theories of suicide, implications for public health intervention, and future directions for using social media to reduce suicide.

The notion of using ML approaches to make interpretations of large data has been explored previously. Poulin et al. demonstrated the capacity for an algorithm to identify suicide risk by analyzing clinical notes [48]. Provided the clinical context of the notes, which included specific references to suicidality, their findings may be somewhat expected. Another step beyond, then, is the analysis of data, which were not intended for a professional audience and to identify the user at suicidal risk. It is believed that text in social media includes technical jargon or official diagnoses indicative of suicidal risk. Only a handful of studies have examined measurement of suicidality in social media data using a variety of methodological approaches. One showed that simply tweeting the phrases “want to commit suicide” or “want to die” were predictive of suicidal ideation and behavior [49]. At least two studies have employed machine-learning algorithms to assess suicide risk. One compared level of agreement between humans and algorithms on three categories of suicide risk, finding rates of agreement between 76–80% [22]. Jashinsky and colleagues compared Twitter derived assessment of suicide risk with rates of suicide from the Centers for Disease Control and Prevention (CDC) and showed the correlation between their method and the actual suicidal rates by state across the US [19]. Our study advanced this line of research by validating Twitter data against already validated measures of suicidality at the individual level. This is an important step forward, because the most effective interventions target individuals who are mostly at risk, ideally with an approach that is tailored to their specific needs [50]. Efforts to target individuals at risk would not need to discuss suicide explicitly, but could simply be a directed tweet mentioning, “In moments of crisis, 1-800-273-TALK is a great resource staffed with trained professionals who care.” It is possible that an individual may feel upset at being targeted with such message. However, this is only a speculation and future research could explore tolerance for such messaging approaches and ultimately inform health communication strategies that are nonobtrusive, yet effective.

The fact that tweets including themes of achievement differentiated respondents so well may at first seem surprising in light of theories of suicide. The interpersonal theory of suicide predicts that suicide is most likely when a sense of thwarted belonging and burdensomeness are coupled with an acquired capability for suicide through repeated exposure to painful and provocative experiences [33]. Other prominent theories focus on hopelessness [51] and escape from the self [52]; none include achievement as a key theoretical component. Further, previous research on achievement as a predictor of suicide generally shows no association while controlling depressive symptoms and other common covariates [53–56].

However, it is likely that achievement helps us to rule out suicide rather than to rule it in. Our algorithm appeared to go through two major steps. First, ruling out those who are clearly nonsuicidal (using achievement) and then ruling suicidality in using themes of death and emotional intensity. Traditional assessment of suicide risk has implicitly focused on discerning severe suicidality among populations that often present with thoughts of suicide (eg, individuals seeking treatment for depression or posttraumatic stress disorder); traditional

approaches do not typically assess which variables rule out suicide. Our attempt to measure suicidality using social media data in a nonclinical population is distinct from typical methods because it does not ask people to report on symptoms, instead the algorithm monitors a broadcast of comments intended to be shared with anyone who will listen. Hence, achievement likely emerged as a strong differentiating factor because the forward thinking and optimistic nature of achievement is antithetical to suicide. Future research should continue to explore whether a similar “rule out” followed by “ruling in” approach occurs in other machine learning algorithms of social media data.

Limitations and Strengths

Our study has a number of limitations as well as strengths. First, although self-report measures have an element of socially desirable responding influencing scores, it is possible that social desirability may also play a role in Twitter data. However, when themes predictive of suicide emerge in social media, and thus go against the typical scripts of social media chatter, they could represent a major cry for help that may be more informative than other methods of assessing suicidal risk; we propose that future research should explore this issue. We also encourage examination of other forms of online media (eg, Facebook, blogs, etc) because they may serve a slightly different function than Twitter and thus generate different algorithms for detecting suicide risk. Second, as suicide is a rare event, only limited amounts of clinically significant suicidality was analyzed. Although we cross-validated our own sample, we encourage other researchers to replicate our work in other samples to provide even stronger converging evidence of these machine-learning algorithms. We would especially encourage replication using samples recruited via means other than MTurk, since it is possible that MTurk participants are different from the general population of social media users in ways that influenced the themes we observed in our research (eg, themes of achievement). On the other hand, our study is the first to validate machine-learning algorithms in Twitter data against psychometrically validated measures of suicidality. Moreover, our multimodal assessment of suicidality took place within a sample that is known to be more attentive [57] and representative than college student populations [58], where novel research ideas are often tested. Further, our results provide strong evidence that we are reliably able to differentiate those who are clinically significantly suicidal from those who are not.

Public Health Significance

Regarding public health approaches to suicide, Twitter offers an unprecedented stream of data connecting individuals to society; our study suggests that there might be a very tangible way we can leverage this phenomenon to do something beneficial. As we further refine our ability to identify suicide risk in real time, our ability to reduce risk for suicide will increase. This may augment existing programs attempting to reduce suicide. For example, suicide hotlines have staff that wait for individuals in crisis to call in; we may enhance these efforts using social media data to proactively identify those who may benefit from their services. When an individual’s public twitter stream indicates clinically significant suicidality, simple interventions such as sending them a private message directing

them to 1-800-273-TALK is almost effortless to do, but may have a significant impact. Simple interventions that foster belongingness or connect people to reach-out-and-talk-to-someone resources are likely to help anyone virtually. An important study showed that simply sending follow-up letters to individuals who had been previously hospitalized for suicide or depression reduced the rate of subsequent suicide compared to those who received no such contact[59]. We believe that expanding our portfolio of approaches to include surveillance of social media in order to identify and prevent suicide across the entire population of those who use social media has the

potential to substantially reduce the incidence of suicide in the US.

The White House has indicated that suicide prevention is a top priority and has funded a number of initiatives attempting to reduce suicide [60]. However, many attempts to reduce suicide are marked by good intentions but lack a strong empirical base and reach only a limited number of people. In order to extend our reach in a way that can commensurate with the problem of suicide, we need to move beyond status quo approaches that wait for people to seek treatment when they are in deep distress and instead seek them out before they reach the point of crisis.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. 2015. Fatal injury data URL: <http://www.cdc.gov/injury/wisqars/fatal.html> [accessed 2016-04-22] [WebCite Cache ID 6gxZ6RHuC]
2. World Health Organization. Prevention suicide: a global imperative. Geneva, Switzerland: World Health Organization; 2014.
3. Gaynes BN, West SL, Ford CA, Frame P, Klein J, Lohr KN, U.S. Preventive Services Task Force. Screening for suicide risk in adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2004 May 18;140(10):822-835. [Medline: [15148072](#)]
4. Hayashi H, Asaga A, Sakudoh M, Hoshino S, Katsuta S, Akine Y. [Linac based radiosurgery; a technical report]. *No Shinkei Geka* 1992 Jul;20(7):769-773. [Medline: [1630568](#)]
5. Scott M, Wilcox H, Huo Y, Turner JB, Fisher P, Shaffer D. School-based screening for suicide risk: balancing costs and benefits. *Am J Public Health* 2010 Sep;100(9):1648-1652. [doi: [10.2105/AJPH.2009.175224](#)] [Medline: [20634467](#)]
6. Peña JB, Caine ED. Screening as an approach for adolescent suicide prevention. *Suicide Life Threat Behav* 2006 Dec;36(6):614-637. [doi: [10.1521/suli.2006.36.6.614](#)] [Medline: [17250467](#)]
7. Schwartz HA, Ungar LH. Data-Driven Content Analysis of Social Media: A Systematic Overview of Automated Methods. *The ANNALS of the American Academy of Political and Social Science* 2015 Apr 09;659(1):78-94. [doi: [10.1177/0002716215569197](#)]
8. Aslam AA, Tsou M, Spitzberg BH, An L, Gawron JM, Gupta DK, et al. The reliability of tweets as a supplementary method of seasonal influenza surveillance. *J Med Internet Res* 2014;16(11):e250 [FREE Full text] [doi: [10.2196/jmir.3532](#)] [Medline: [25406040](#)]
9. Paul M, Dredze M. You are what you tweet: analyzing Twitter for public health. 2011 Presented at: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media; 2011; Barcelona, Spain.
10. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, Seligman Martin E P. Psychological language on Twitter predicts county-level heart disease mortality. *Psychol Sci* 2015 Feb;26(2):159-169 [FREE Full text] [doi: [10.1177/0956797614557867](#)] [Medline: [25605707](#)]
11. West JH, Hall PC, Hanson CL, Prier K, Giraud-Carrier C, Neeley ES, et al. Temporal variability of problem drinking on Twitter. *OJPM* 2012;02(01):43-48. [doi: [10.4236/ojpm.2012.21007](#)]
12. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. *J Med Internet Res* 2013;15(9):e189 [FREE Full text] [doi: [10.2196/jmir.2741](#)] [Medline: [24014109](#)]
13. Hanson CL, Burton SH, Giraud-Carrier C, West JH, Barnes MD, Hansen B. Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. *J Med Internet Res* 2013;15(4):e62 [FREE Full text] [doi: [10.2196/jmir.2503](#)] [Medline: [23594933](#)]
14. Prier K, Smith M, Giraud-Carrier C, Hanson C. Identifying health related topics on twitter: An exploration of tobacco-related tweets as a test topic. 2011 Presented at: Proceedings of the 4th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction; March 29-31, 2011; College Park, MD.
15. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009;11(1):e11 [FREE Full text] [doi: [10.2196/jmir.1157](#)] [Medline: [19329408](#)]
16. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. *Am J Prev Med* 2011 May;40(5 Suppl 2):S154-S158. [doi: [10.1016/j.amepre.2011.02.006](#)] [Medline: [21521589](#)]
17. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Communication* 2015 Jul;71:10-49. [doi: [10.1016/j.specom.2015.03.004](#)]

18. De Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. 2013 Presented at: Proceedings of the 5th Annual ACM Web Science Conference; March 2-4, 2013; Paris, France.
19. Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, et al. Tracking suicide risk factors through Twitter in the US. *Crisis* 2014;35(1):51-59. [doi: [10.1027/0227-5910/a000234](https://doi.org/10.1027/0227-5910/a000234)] [Medline: [24121153](https://pubmed.ncbi.nlm.nih.gov/24121153/)]
20. Coppersmith G, Dredze M, Harman C, Hollingshead K, Mitchell M. CLPsych 2015 shared task: Depression and PTSD on Twitter. 2015 Presented at: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology; May 31-June 5, 2015; Denver, Colorado.
21. Huang X, Zhang L, Chiu D, Liu T, Li X, Zhu T. Detecting suicidal ideation in Chinese microblogs with psychological lexicons. 2014 Presented at: Proceedings of the 11th International Conference on Ubiquitous Intelligence and Computing and 11th International Conference on Autonomic and Trusted Computing and 14th International Conference on Scalable Computing and Communications; December 9-12, 2014; Bali, Indonesia.
22. O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on Twitter. *Internet Interventions* 2015 May;2(2):183-188. [doi: [10.1016/j.invent.2015.03.005](https://doi.org/10.1016/j.invent.2015.03.005)]
23. Guan L, Hao B, Cheng Q, Yip PS, Zhu T. Identifying Chinese Microblog Users With High Suicide Probability Using Internet-Based Profile and Linguistic Features: Classification Model. *JMIR Mental Health* 2015 May 12;2(2):e17. [doi: [10.2196/mental.4227](https://doi.org/10.2196/mental.4227)]
24. Sueki H. The association of suicide-related Twitter use with suicidal behaviour: a cross-sectional study of young internet users in Japan. *J Affect Disord* 2015 Jan 1;170:155-160. [doi: [10.1016/j.jad.2014.08.047](https://doi.org/10.1016/j.jad.2014.08.047)] [Medline: [25240843](https://pubmed.ncbi.nlm.nih.gov/25240843/)]
25. Kumar M, Dredze M, Coppersmith G, De CM. Detecting changes in suicide content manifested in social media following celebrity suicides. 2015 Presented at: Proceedings of the 26th ACM Conference on Hypertext & Social Media; 2015; KalKanli, Cyprus.
26. Thompson P, Bryan C, Poulin C. Predicting military and veteran suicide risk: Cultural aspects. In: Predicting military and veteran suicide risk: Cultural aspects. 2014 Presented at: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology; June 27, 2014; Baltimore, Maryland.
27. Coppersmith G, Dredze M, Harman C. Quantifying Mental Health Signals on Twitter. 2014 Presented at: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology; June 27, 2014; Baltimore, Maryland.
28. Schwartz H, Eichstaedt J, Kern M, Park G, Sap M, Stillwell D, et al. Towards Assessing Changes in Degree of Depression through Facebook. 2014 Presented at: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology; June 27, 2014; Baltimore, Maryland.
29. Mitchell M, Hollingshead K, Coppersmith G. Quantifying the Language of Schizophrenia in Social Media. 2015 Presented at: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology; May 31-June 5, 2015; Denver Colorado.
30. Pedersen T. Screening Twitter Users for Depression and PTSD with Lexical Decision Lists. 2015 Presented at: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology; May 31-June 5, 2015; Denver, Colorado.
31. Robinson J, Cox G, Bailey E, Hetrick S, Rodrigues M, Fisher S, et al. Social media and suicide prevention: a systematic review. *Early Interv Psychiatry* 2016 Apr;10(2):103-121. [doi: [10.1111/eip.12229](https://doi.org/10.1111/eip.12229)] [Medline: [25702826](https://pubmed.ncbi.nlm.nih.gov/25702826/)]
32. Joiner TE, Pfaff JJ, Acres JG. A brief screening tool for suicidal symptoms in adolescents and young adults in general health settings: reliability and validity data from the Australian National General Practice Youth Suicide Prevention Project. *Behav Res Ther* 2002 Apr;40(4):471-481. [Medline: [12008659](https://pubmed.ncbi.nlm.nih.gov/12008659/)]
33. Van Orden Kimberly A, Witte TK, Cukrowicz KC, Braithwaite SR, Selby EA, Joiner TE. The interpersonal theory of suicide. *Psychol Rev* 2010 Apr;117(2):575-600 [FREE Full text] [doi: [10.1037/a0018697](https://doi.org/10.1037/a0018697)] [Medline: [20438238](https://pubmed.ncbi.nlm.nih.gov/20438238/)]
34. Van Orden Kimberly A, Witte TK, Gordon KH, Bender TW, Joiner TE. Suicidal desire and the capability for suicide: tests of the interpersonal-psychological theory of suicidal behavior among adults. *J Consult Clin Psychol* 2008 Feb;76(1):72-83. [doi: [10.1037/0022-006X.76.1.72](https://doi.org/10.1037/0022-006X.76.1.72)] [Medline: [18229985](https://pubmed.ncbi.nlm.nih.gov/18229985/)]
35. Peer E, Vosgerau J, Acquisti A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behav Res Methods* 2014 Dec;46(4):1023-1031. [doi: [10.3758/s13428-013-0434-y](https://doi.org/10.3758/s13428-013-0434-y)] [Medline: [24356996](https://pubmed.ncbi.nlm.nih.gov/24356996/)]
36. Pennebaker J, Boyd R, Jordan K, Blackburn K. The development and psychometric properties of LIWC. 2015. URL: <http://www.liwc.net/LIWC2007LanguageManual.pdf> [accessed 2016-04-22] [WebCite Cache ID 6gxWUzbZrq]
37. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 2011 Sep 30;333(6051):1878-1881 [FREE Full text] [doi: [10.1126/science.1202775](https://doi.org/10.1126/science.1202775)] [Medline: [21960633](https://pubmed.ncbi.nlm.nih.gov/21960633/)]
38. De Choudhury M, Counts S, Horvitz E. Major life changes and behavioral markers in social media: case of childbirth. 2013 Presented at: Proceedings of the Conference on Computer Supported Cooperative Work; February 23-27, 2013; San Antonio, Texas.
39. Stirman SW, Pennebaker JW. Word use in the poetry of suicidal and nonsuicidal poets. *Psychosom Med* 2001;63(4):517-522. [Medline: [11485104](https://pubmed.ncbi.nlm.nih.gov/11485104/)]
40. Garcia-Caballero A, Jimenez J, Fernandez-Cabana M, Garcia-Lado I. Last Words: An LIWC Analysis of Suicide Notes from Spain. *Eur Psychiat* 2012;27.
41. Fernández-Cabana M, García-Caballero A, Alves-Pérez MT, García-García MJ, Mateos R. Suicidal traits in Marilyn Monroe's Fragments: an LIWC analysis. *Crisis* 2013;34(2):124-130. [doi: [10.1027/0227-5910/a000183](https://doi.org/10.1027/0227-5910/a000183)] [Medline: [23261910](https://pubmed.ncbi.nlm.nih.gov/23261910/)]

42. Homan C, Johar R, Liu T, Lytle M, Silenzio V, Ovesdotter AC. Toward macro-insights for suicide prevention: Analyzing fine-grained distress at scale. 2014 Presented at: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology; June 22-27, 2014; Baltimore, Maryland.
43. Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. 2015 Presented at: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology; May 31-June 5, 2015; Denver, Colorado.
44. Kang R, Brown S, Dabbish L, Kiessler S. Privacy attitudes of Mechanical Turk workers and the U.S. public. 2014 Presented at: Proceedings of the Symposium on Usable Privacy and Security (SOUPS); July 9-11, 2014; Melo Park, California.
45. Silenzio Vincent M B, Duberstein PR, Tang W, Lu N, Tu X, Homan CM. Connecting the invisible dots: reaching lesbian, gay, and bisexual adolescents and young adults at risk for suicide through online social networks. *Soc Sci Med* 2009 Aug;69(3):469-474 [FREE Full text] [doi: [10.1016/j.socscimed.2009.05.029](https://doi.org/10.1016/j.socscimed.2009.05.029)] [Medline: [19540641](https://pubmed.ncbi.nlm.nih.gov/19540641/)]
46. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *JMLR* 2011;12:2825-2830.
47. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. New York: Chapman & Hall; 1984.
48. Poulin C, Shiner B, Thompson P, Vepstas L, Young-Xu Y, Goertzel B, et al. Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS ONE* 2014;9(1):e85733. [doi: [10.1371/journal.pone.0091602](https://doi.org/10.1371/journal.pone.0091602)]
49. Sueki H. The association of suicide-related Twitter use with suicidal behaviour: a cross-sectional study of young internet users in Japan. *J Affect Disord* 2015 Jan 1;170:155-160. [doi: [10.1016/j.jad.2014.08.047](https://doi.org/10.1016/j.jad.2014.08.047)] [Medline: [25240843](https://pubmed.ncbi.nlm.nih.gov/25240843/)]
50. Coie JD, Watt NF, West SG, Hawkins JD, Asarnow JR, Markman HJ, et al. The science of prevention. A conceptual framework and some directions for a national research program. *Am Psychol* 1993 Oct;48(10):1013-1022. [Medline: [8256874](https://pubmed.ncbi.nlm.nih.gov/8256874/)]
51. Beck AT, Brown GK, Steer RA, Kuyken W, Grisham J. Psychometric properties of the Beck Self-Esteem Scales. *Behav Res Ther* 2001 Jan;39(1):115-124. [Medline: [11125720](https://pubmed.ncbi.nlm.nih.gov/11125720/)]
52. Baumeister RF. Suicide as escape from self. *Psychol Rev* 1990;97(1):90-113. [doi: [10.1037//0033-295X.97.1.90](https://doi.org/10.1037//0033-295X.97.1.90)]
53. Canetto SS, Lester D. Love and achievement motives in women's and men's suicide notes. *J Psychol* 2002 Sep;136(5):573-576. [doi: [10.1080/00223980209605552](https://doi.org/10.1080/00223980209605552)] [Medline: [12431040](https://pubmed.ncbi.nlm.nih.gov/12431040/)]
54. Hull-Blanks EE, Kerr BA, Robinson Kurpius Sharon E. Risk factors of suicidal ideations and attempts in talented, at-risk girls. *Suicide Life Threat Behav* 2004;34(3):267-276. [doi: [10.1521/suli.34.3.267.42782](https://doi.org/10.1521/suli.34.3.267.42782)] [Medline: [15385181](https://pubmed.ncbi.nlm.nih.gov/15385181/)]
55. Klibert J, Langhinrichsen-Rohling J, Luna A, Robichaux M. Suicide proneness in college students: relationships with gender, procrastination, and achievement motivation. *Death Stud* 2011 Aug;35(7):625-645. [Medline: [24501841](https://pubmed.ncbi.nlm.nih.gov/24501841/)]
56. Lewis SA, Johnson J, Cohen P, Garcia M, Velez CN. Attempted suicide in youth: its relationship to school achievement, educational goals, and socioeconomic status. *J Abnorm Child Psychol* 1988 Aug;16(4):459-471. [Medline: [3221034](https://pubmed.ncbi.nlm.nih.gov/3221034/)]
57. Hauser DJ, Schwarz N. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behav Res Methods* 2016 Mar;48(1):400-407. [doi: [10.3758/s13428-015-0578-z](https://doi.org/10.3758/s13428-015-0578-z)] [Medline: [25761395](https://pubmed.ncbi.nlm.nih.gov/25761395/)]
58. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 2011 Feb 03;6(1):3-5. [doi: [10.1177/1745691610393980](https://doi.org/10.1177/1745691610393980)]
59. Motto JA, Bostrom AG. A randomized controlled trial of postcrisis suicide prevention. *Psychiatr Serv* 2001 Jun;52(6):828-833. [Medline: [11376235](https://pubmed.ncbi.nlm.nih.gov/11376235/)]
60. Clay R. Monitor on Psychology. Suicide prevention is a top White House priority URL: <http://www.apa.org/monitor/2014/11/upfront-suicide.aspx> [accessed 2016-04-22] [WebCite Cache ID 6gxYxNd1S]

Edited by P Bamidis; submitted 16.06.15; peer-reviewed by G Coppersmith, B Spitzberg, L Litman, MH Tsou, S Konstantinidis; comments to author 20.07.15; revised version received 04.02.16; accepted 25.02.16; published 16.05.16

Please cite as:

Braithwaite SR, Giraud-Carrier C, West J, Barnes MD, Hanson CL

Validating Machine Learning Algorithms for Twitter Data Against Established Measures of Suicidality

JMIR Mental Health 2016;3(2):e21

URL: <http://mental.jmir.org/2016/2/e21/>

doi: [10.2196/mental.4822](https://doi.org/10.2196/mental.4822)

PMID: [27185366](https://pubmed.ncbi.nlm.nih.gov/27185366/)

©Scott R. Braithwaite, Christophe Giraud-Carrier, Josh West, Michael D. Barnes, Carl Lee Hanson. Originally published in *JMIR Mental Health* (<http://mental.jmir.org>), 16.05.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Mental Health*, is properly cited. The complete

bibliographic information, a link to the original publication on <http://mental.jmir.org/>, as well as this copyright and license information must be included.