Detecting Low Self-Esteem in Youths from Web Search Data

Anis Zaman
Department of Computer Science
University of Rochester
Rochester, NY
azaman2@cs.rochester.edu

Rupam Acharyya
Department of Computer Science
University of Rochester
Rochester, NY
racharyy@cs.rochester.edu

Henry Kautz
Department of Computer Science
University of Rochester
Rochester, NY
kautz@cs.rochester.edu

Vincent Silenzio
School of Medicine
University of Rochester
Rochester, NY
v.m.silenzio@rochester.edu

ABSTRACT

Online behavior leaves a digital footprint that can be analyzed to reveal our cognitive and psychological state through time. Recognizing these subtle cues can help identify different aspects of mental health, such as low self-esteem, depression, and anxiety. Google's web search engine, used daily by millions of people, logs every search query made by a user, which is accessible through a platform called Google Takeout. Previous researchers have made efforts to detect and predict behaviors associated with depression and anxiety from web data, but only at a population level. This paper fills in the gap of looking into signs of low self-esteem, a condition that work in a vicious cycle with depression and anxiety, at an individual level by looking into Google search history data. We target college students, a population prone to depression, anxiety, and low self-esteem, and ask to take mental health assessment survey along with their individual search history. Textual analysis show that search logs contain strong signals that can identify individuals with current low self-esteem. For example, participants with low self-esteem have fewer searches pertaining to family, friend, and money attributes; and we also observed differences in the search category distribution, over time, when compared with individuals with moderate to high self-esteem. Using these markers we were able to build a probabilistic classifier that can identify low self-esteem conditions, based on search history, with an average F1 score of 0.86.

KEYWORDS

low self-esteem, mental health, youths, college students, individual search logs

ACM Reference Format:

Anis Zaman, Rupam Acharyya, Henry Kautz, and Vincent Silenzio. 2019. Detecting Low Self-Esteem in Youths from Web Search Data. In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3308558.3313557

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313557

1 INTRODUCTION

Mental health can be viewed as a complex interaction between biological, genetic, social, and environmental factors. Due to the multifaceted nature and stigma of mental health, people are likely to rely on online services to investigate their problems in the World Wide Web [24, 50]. Google search is one such platform that is used by millions of people everyday and approximately 3.5 billion searches are made every day with 40,000 queries per second[1]. Hence, online searches may leave digital footprints that can be analyzed to reveal our cognitive and emotional state through time. Detecting these subtle cues can help therapists and others better recognize different components of mental health such as anxiety, stress, depression, and low self-esteem. Due to the complex nature of mental health and various factors associated to it, identifying individuals who are at risk or need help has been a challenge in the mental health research community [4, 8, 45]. For example, among college students, low self-esteem and stress are conditions that are more likely to correlate with suicidal ideation [51].

Our work brings significant new approaches and insights to the table: (i) Self-esteem has been rarely addressed in the AI literature before as part of the discussion of depression, and our approach provides a more expansive parsing of the conceptual frameworks in which constructs such as "depression," "self-esteem," and "suicidal ideation" are embedded [48]. (ii) We provide a scalable, lightweight solution that can be deployed at the point-of-care to use immediately available search history data to provide instantly actionable therapeutic targets for clinicians. (iii) Our work opens a way for clinicians to obtain access to patients' internal mental health states and traits, with high fidelity and accuracy far beyond the limits of traditional self-reported data: we are, in fact, now incorporating the approach into planned clinical trials.

Researchers have been using social media data to identify and predict different mental health conditions and associated factors [1, 6, 15, 15, 40]. Twitter, Facebook, Reddit, Instagram, and Weibo are some of the common data sources exploited by researchers to explore signs of depression, anxiety, suicidal ideation, and other related conditions.

With only a few exceptions [11, 12], these prior studies have not had access to ground-truth data about the subjects' mental health as assessed by a validated clinical instrument. While they make

[1] http://www.internetlivestats.com/google-search-statistics/

claims about depression at a population or individual level, they cannot justify those claims in a manner that would be actionable by a healthcare professional.

Another limitation of most prior work is that data is only gathered from users of particular social media sites. Because users self-select for particular sites, there may well be confounding factors that influence both the users' mental health and their online behavior. For example, people who routinely broadcast their thoughts on Twitter may have different ways of talking about depression than more private individuals. A further confounding factor is that users may choose to self-censor when posting to social media [22].

In this paper we investigate observable textual signatures related to current low self-esteem condition using both Google search histories and answers to a validated mental health assessment survey. For search, as opposed to use of social media, users do not self-select or face the same temptations to self-censor. Another advantage of search data is that the nature and volume of search queries give temporal behavioral information about a user. While some prior researchers have analyzed search logs for large aggregate populations using Google Trends [2, 55], applying their results to individuals may result in ecological errors. In addition, Google Trends data cannot be used to understand or predict *individual* behavior.

In summary, our dataset and analyses are unique in that they (i) include gold-standard individual psychological assessment data; (ii) use data from a form of online behavior (search) that engages all internet users — that is, 89% of the US adult population [46]; and (iii) employ a data-collection methodology (described below) that is scalable to hundreds or thousands of subjects while preserving users' privacy. Psycholinguistics analysis identified variables capable of differentiating people with low self-esteem. When compared with someone with moderate to high self-esteem, we observed noticeable differences in the type of searches made by individuals with low self-esteem condition. We demonstrated that these identifiers can be used to build statistical models that do better than chance for classifying users with low self-esteem condition using historical search data.

2 RELATED WORK

Researchers have used data from social media, public forums, blogs, and search engine logs to explore interaction of various aspects of mental health, including depression, anxiety, suicidal ideation, mood, and stress. De Choudhury provides an excellent overview of the role of social media in mental health research [11]. Her own previous work showed that lexicons in Tweets contain strong signals for predicting risk of major depressive disorder [14], and she recently identified linguistic markers to identify Twitter users likely to develop future suicidal ideation [15]. In 2018, Liu et al. examined statuses of Weibo users and demonstrated that users exposed to domestic violence showed increase in depressive symptoms, higher suicide risk, and decrease life satisfaction when compared against control group [30]. Other researchers have shown that Tweets contains meaningful clues for various mental health concerns including Adderall use [23], insomnia [26], and depression [10–14, 16].

Perhaps the dataset closest to ours which has been explored by researchers in the past is search logs from Google Trends [19]. Researchers have utilized Google Trends data to demonstrate deeper insight about population behavior and health related phenomena to build predictive models around mental and physical health assessment. For example, surveillance of influenza outbreaks [9, 18]; monitoring toxicological outbreaks [54]; identifying seasonality in seeking mental health [2]; monitoring suicide related terms to help predict rates of self-injury and death within the population [31]; and effects of geographic location on depression related searches [52] are some of the areas that researchers have explored. Nuti et al. provides a through review of the usage of Google Trends in the health care domain [33]. Sueki et al. investigated variations in the volume of Google searches about suicide and depression in Japan and found that the monthly search volumes for "suicide" and "suicide method" were not significantly correlated with the monthly suicide rate, but searches for "depression" were [49]. Yang et al. replicated this associations between the volume of Google search terms and the monthly suicide rate in Taipei City [53]. Gunn and Lester found that suicide rate for Americans was positively correlated with the search volume for search terms such as "commit suicide", "how to commit suicide", and "suicide prevention" [20]. They showed that monitoring internet searches for suicide related terms may provide a faster way to capture, in the moment, possible trends in mental health conditions such as depression, anxiety, and suicide. Song et al. showed that suicide and stress related searches are positively associated with suicide rates in Korea [47]. Hagihara et al. demonstrated that Google search volume for terms associated with hydrogen sulfide in Japan was correlated with the monthly suicide rate for people in their 20's and 30's [21]. Ayers et al. showed that there was a significant increase in suicide related terms in Google for the 19 days following the release of a popular Netflix show *13 Reasons Why*, which narrates the suicide of a fictional teen [3]. However, Page et al. found that internet search trends related to suicide terms are not necessarily straightforward indicators of the degree of suicide behavior in Australia [36]. It is debatable whether any query preceded an actual suicide attempt; but it is quite evident that search logs monitoring can be used to access understanding of different mental health conditions and capture any potential signals of low self-esteem, depression, anxiety, self harm risk, etc.

All of the work above was performed on search logs aggregated over a large population. By contrast, our work is the first to investigate search history data at an *individual level* for signs of low self-esteem.

3 DATA

Data for this study were collected through a 2-month on-campus recruitment to investigate how online services such as Google search can be used to detect early symptoms of suicidal ideation, depression, anxiety, low self-esteem, and lack of social supports. As described below, we developed a Human Subjects Review Board approved cloud-based application to acquire consent from subjects, download and anonymize their search histories, and link anonymized histories to corresponding mental health survey answers.

3.1 Participants

In total 108 individuals, 40 male and 68 female, volunteered to participate. Subjects were recruited through advertisements. To limit any participation bias, the information on the advertisement was generic and had no mention of mental health related phenomena because some participants with certain personality traits can avoid recruitment.

A step in the data collection involves subjects completing an online written questionnaire about mental health, anxiety, suicidal ideation, and self-esteem. We call this the *Promote Health* survey. In addition to academic and demographic information, the survey incorporates questions from several standard instruments such as the depression module of the Patient Health Questionnaire (PHQ-9) [29] and the Rosenberg self-esteem scale [41, 43]. PHQ-9 is a common psychological instruments that has been used by previous researchers for depression related screening while mining social media data [27, 56]. Rosenberg self-esteem scale has been used to study usage of social media and its' association with low self-esteem among undergraduate students [39].

3.2 Recruitment Procedure

Interested participants were required to be at least 18 years old, able to read and write in English, and have an active Google services account. Qualified subjects were informed about the purpose, procedure, and goal of the study; then they reviewed and signed a statement of informed consent explaining the Human Subjects Committee approval. The study had 100% participation rate, meaning none of the 108 individuals decided to opt out of the study after the study was explained to them.

There are two parts to the data collection process. First, study recruiter asks participants to type their Google login credentials into a web-based app running on a tablet. Second, study participants were asked to take the *Promote Health* survey, a brief self assessment online questionnaire. There are 38 freshmen, 18 sophomore, 17 juniors, and 19 seniors in our dataset. In addition, there were 12 graduate students and 4 fifth year undergraduates. Participants' mean GPA was 3.51. Note that we did not make use of these additional data in the analyses presented in this paper, but that it may help explain the surprising level of low-self esteem described below.

3.3 Google Search History

Anyone with a Google account can freely access his or her entire personal data, at any time, through a platform called Google Takeout[2]. Using Google Oauth authentication protocol, participants authorize our search download application to access users Google related data. Although data about many forms of Google services can be accessed through this portal, this study uses only data from Google search. None of the other potential sources of data are accessed or stored at any time. Individuals who did not have Google search history tracking enabled were excluded from our analysis.

Once a user authorizes the application to obtain a copy of the search history, a one-time log in authorization is used to automatically transfer the search data from user's google drive to a secured server in the cloud. The oauth token that our application uses to access a user's data is valid only for 24 hours.

Before any data is transferred from "the cloud" and given to the research team to store as part of this study, the data is de-identified via Google's Cloud Data Loss Prevention (DLP) API. The API scans for personal identifiers such as names, addresses, phone numbers, etc., that can potentially identify individual study participants, and this information is automatically removed from the search history data prior to being saved as research data.

Once the search history data is successfully de-identified and transferred to the research team, the one-time log in authorization is permanently deleted. The research team is no longer able to use this authorization to access any participant's online data again. Fig. 1 presents the entire data collection process.

[2]https://takeout.google.com/settings/takeout

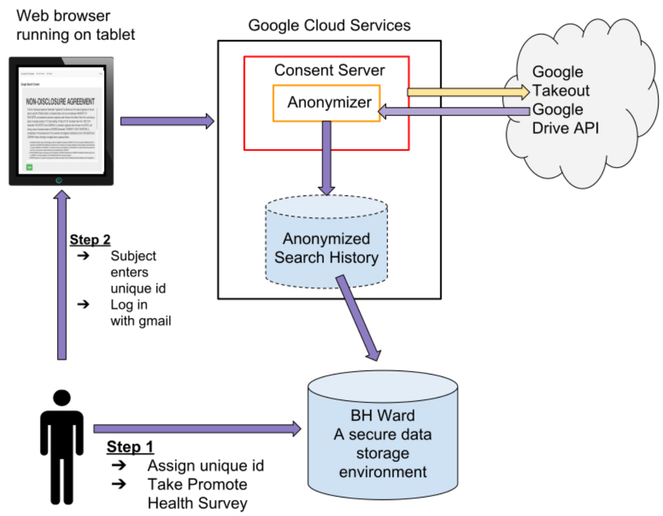

Figure 1: Data Collection Process.

Low Self-esteem (LS)	No Low Self-esteem (NLS)	
27	28	Analysis or Training set
21	19	Test set
5	8	No search data. Excluded from dataset
Total 108 subjects		

Table 1: Search dataset breakdown

3.4 Promote Health

Promote Health is a computer assisted survey instrument (CASI), designed at a fifth grade reading level and includes questions about age, academic concentration, gpa, race, gender, mental health (PHQ-9), Interpersonal violence (IPV), sexual health, self-esteem, etc. Users did not write their names on the survey at anytime, but instead entered a random ID number that was used to link their responses to their search history data. No records connecting subject names to ID numbers was maintained.

3.5 Data Description

A section in the survey asks series of questions for evaluating individual self-esteem. The answer to the self-esteem questions were used to compute the Rosenberg Self-Esteem Scale(RSES). 53 individuals had RSES score less than 12 (on a scale of 0-30, people with average to high self-esteem have score ≥ 15), indicating low self-esteem(LS). On the contrary, 55 individuals had score greater than or equal to 12 indicating no sign of low self-esteem(NLS). However, there were 13 individuals who participated but did not have any search data and were excluded from any analysis.

Before any analysis and modeling, we separate the dataset into two parts, namely **analysis set** and **test set**. This is illustrated in Table 1. For the self-esteem analysis, 27 participants had LS and 28 had NLS condition. The test set had 19 subjects with NLS and 21 with LS condition.

Through the *Promote Health*, we measured the current RSES score. Low self-esteem is one of the vulnerable factors that works in a vicious cycle with depression, anxiety, and suicide ideation [5]. However, low self-esteem state and any specific mental health condition may not necessarily occur at the same time but can have indirect effect on the quality of life.

In our survey responses, we observed that low self-esteem condition stood out strongly than depression in the responses. The average gpa of the participants is 3.51, which is quite high. A potential explanation can be "imposter syndrome" – where high performing students do not feel they are as smart as their colleagues. Young college undergraduates are more prone to self-criticism, social anxiety, achievement pressures, and depressive tendencies [28].

4 OBSERVATION ON GOOGLE SEARCH LOGS

Google search phrases tend to be short, recorded in the present and are good indicators of what may be going through an individual's mind at the moment. The context of a search phrase can be anything. We are interested in identifying the categories of the searches and uncover any linguistic signals in the search phrases that can help distinguish subjects with LS and NLS conditions. In the following sections, all the search logs analysis is carried out on the **analysis set**, see Fig 1.

4.1 Variance of Search Categories

Using the content classification feature of the Google Cloud NLP API[3] we were able to classify the search queries of every participating subjects. The API returns a hierarchical category label for every search query. We treat the broad category as the 'category' label for the query. For instance, for a query *q*, if the label from the API is '/Arts & Entertainment/Humor/Funny Pictures & Videos', we consider 'Arts & Entertainment' as the category for *q*.

In total, there are 27 such categories. We observed that, at any time of the day, individuals with LS condition consistently searched more on topics related to 'Sports', 'Health', 'Finance', 'News', 'Books & Literature', 'Reference', and 'Law & Government'. This is illustrated in Fig. 2. The 'Reference' category constitutes mostly of topics associated to humanities, language resources, geographic references, and general references such as concepts from Encyclopedias, Wikipedias, etc. One explanation for such high prevalence

[3]https://cloud.google.com/natural-language/docs/classifying-textt

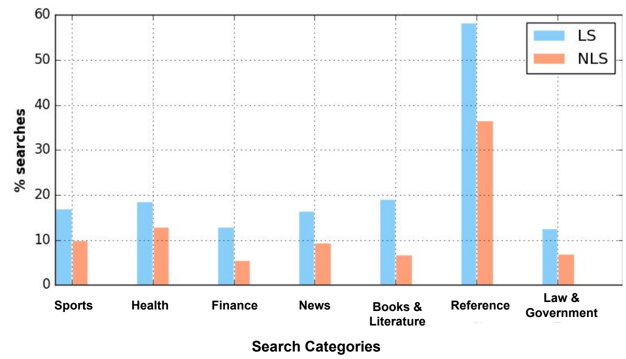

Figure 2: Search categories that were more frequent in LS population

of searches from this category can be that students with low self-esteem condition are more reliant on search engines to do their daily school works compared to the population that does not exhibit low self-esteem.

We also observed that individuals with current LS, on average, search more about certain topics than students with NLS condition during late hours (1AM - 5AM). For instance, students with LS condition made more searches pertaining to categories such as 'Reference', 'Jobs & Education', 'Books & Literature', and 'Health' than students with NLS condition. This increased late night activity disrupts normal sleep cycle which is reported to be one of the core symptoms of depression and can eventually lead to poor quality of mental health and life in general [34].

Context(i.e.category of a search) and frequency of searches on Google vary from person to person. Over the entire search history, we noticed observable differences in distribution of search categories between the LS and NLS population. For example, searches pertaining to the '/Arts & Entertainment' category were more frequent for both the groups. However, there are categories such as 'Health', 'Food & Drink', 'Sensitive Subjects', 'References', etc. for which the distribution demonstrate apparent differences.

4.2 Psycho linguistic attribute detection in search logs

At first, it may not be self-evident that query logs can be good indicators of what may be going through an individual's mind at the moment. Traditionally, researchers have done psycho linguistic analysis on social media data to uncover signals of mental health related phenomena. In 2017, Oexle et al. have shown that there is a strong stigma associated to expressing oneself related to mental illness and other related issues [35]. As a defensive mechanism, users may stop generating content on social media platforms [25]. However, when someone searches, there is no fear of any judgment; and what he/she is searching can potentially reveal his/her psychological state at that moment.

LIWC is a text analysis toolkit that outputs the proportion of words in a given text that fall into one or more of over 80 linguistic (such as prepositions, adverbs, first-person singular pronouns, conjunctions), psychological (happy, anger, achievement, etc.), and topical (leisure, money, etc.) categories [37]. Researchers have been

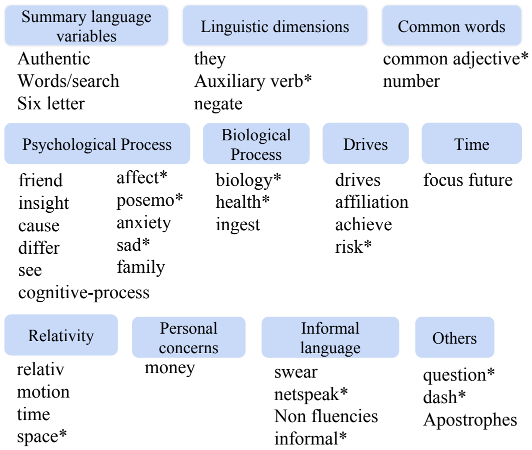

Figure 3: LIWC attributes that were statistically significantly different between the subjects with LS and NLS condition. Concepts with * were more prevalent among subjects with LS condition.

using this tool to study the distribution of analytical thinking, authenticity, emotional tone, social and cognitive process present in a piece of text. Clinicians and social computer scientists do psychometric analysis using LIWC lexicons. Some popular text data analyzed using this lexicon includes social media, online forums, written notes, and self-introductions [42]. We are the first to demonstrate the presence of these linguistic attributes in individual level search data.

For the low self-esteem analysis, our data consists of 27 subjects with LS and 28 subjects with NLS condition from the **analysis set**(see Fig 1). For every subject with LS condition, all the Google search phrases are analyzed using LIWC and the output is a 92 dimension vector. The result can be treated as a table with 27 rows and 92 columns. A similar procedure is repeated for every participants with NLS condition, resulting in a table with 28 rows and 92 columns.

Next, we explore whether there are any statistically significant LIWC categories between the two populations (LS vs NLS). Our sample size is small, assuming normal distribution, we consider the LIWC variables from the LS and NLS groups and do student's t-tests. We found 40 LIWC attributes that were statistically significantly different(with p-value < 0.05) between participants with LS and NLS, as shown in Fig. 3.

In the low self-esteem analysis, the family variable was significantly lower for almost every participant with LS condition than participants with NLS condition. This supports previously established findings where researchers have shown that family support influences self-esteem, which impacts adolescent suicide risk behaviors [32, 44]. Fig. 4 presents a visual representation of the distribution of some of the LIWC attributes (from Fig 3) that are statistically significantly different between LS and NLS.

Search phrases are short and lack syntactic structure compared to other forms of data that are popularly analyzed using LIWC. Yet, we found strong signals in LIWC categories that can differentiate and identify individuals with low self-esteem. Our analysis of search

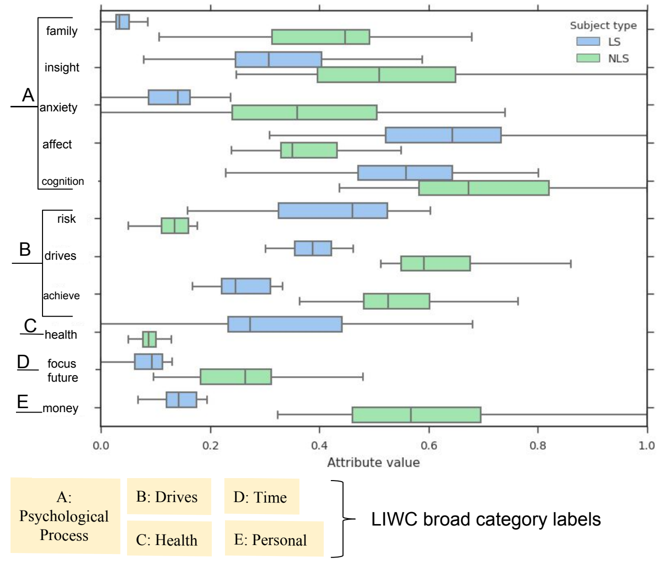

Figure 4: Distribution of some LIWC attributes score between participants with LS and NLS condition. These are some of the attributes from Fig. 3

data strengthens prior findings regarding the association of self esteem with other mental health phenomena [32] .

This indicates that there are some shared lexicons, even in Google search phrases, among individuals with LS condition. For instance Affect, Insight, Cogproc, Space, and Affiliation are some of the linguistic variables that are significantly to differentiate a population between LS and NLS.

5 MODELING LOW SELF-ESTEEM (LS) CONDITION

For the classification task, we ask the question: *Based on the search history, can we identify subjects with low self-esteem?*

We use the **analysis set** from Fig. 1 for selecting features and training our classifiers. Our training set consists of 27 individuals with LS and 28 with NLS condition.

The performance of the trained model is evaluated using a test set. We separated this set from the data prior to the feature selection and the training process. The held out test set consists of 19 subjects with NLS and 21 with LS.

5.1 Feature Selection from Analysis Set

Based on our observation from section 4.1 & 4.2, we consider two types of features for the classification task.

5.1.1 Features based on LIWC:. There were 40 linguistic attributes that were statistically significantly different between the LS and NLS population (see Fig. 3). We treat these LIWC attributes as features for training our models.

5.1.2 Features based on search category distribution: Searching behavior varies from person to person. Knowing the categories of

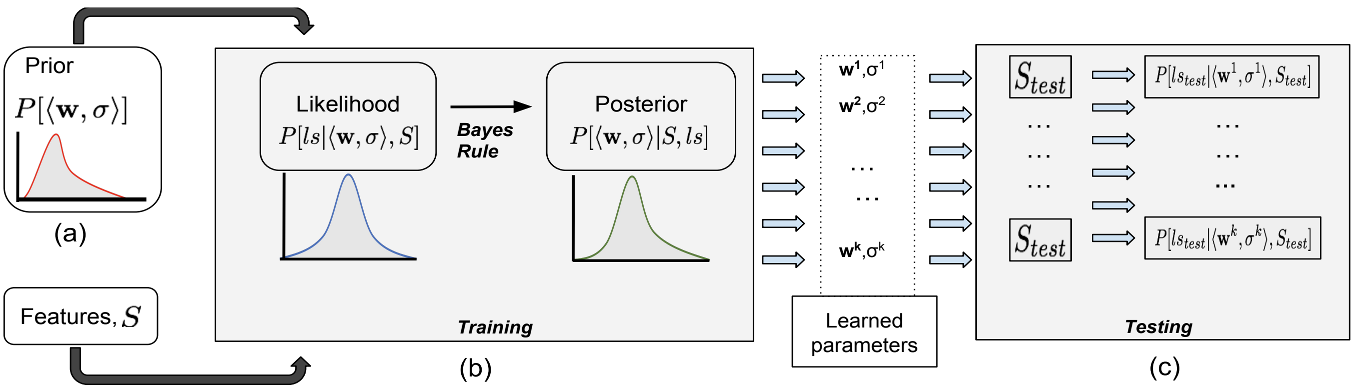

Figure 5: Process depicting learning and testing procedure of the HyBaR model. (a) Prior distribution of the parameters w and σ. Prior of σ is uniform distribution. In case of LIWC based features the prior of w is Laplace distribution defined in (4) and is Gaussian distribution with category similarity matrix as the covariance matrix for the search category based features. (b) Posterior is obtained from the corresponding prior and the likelihood using Bayes Rule. Using Metropolis-Hastings Algorithm several model parameters are sampled from the approximate posterior. (c) S_{test} represents the test data which has been held out prior to any analysis and training. Each of the learned parameters are used individually to classify S_{test}. This helps in model generalization.

searches can reveal insightful patterns in internet activities. Different individuals exhibit different searching behavior. For example, following Fig 3, we found that students with LS condition consistently searched more about certain topics than students with NLS condition. This motivated us to treat the search category distribution, over one's entire search history, as features for our models. Using the Google Cloud NLP API, we labelled search queries into 27 categories. We treat the distribution of these 27 search categories as features and train our classifiers.

5.2 Models

We propose a discriminative model whose parameters are learned using Bayesian approach. We call our model Hybrid Bayesian Regression (HyBaR) because it is a cross-breed between the Bayesian Linear Regression and the Bayesian Logistic Regression [7]. For details about the model, see section 5.2.1.

In addition, we trained a **Logistic Regression(LR)**, a **Support Vector Machine(SVM)** and compare their performance with HyBaR.

5.2.1 ***Hybrid Bayesian Regression (HyBaR):***. HyBaR has advantage over Maximum Likelihood Estimation (MLE) due to the additional prior information. Compared to the maximum a posteriori (MAP) estimation, which outputs a parameter that maximizes the posterior probability for the given data, HyBaR estimates the entire posterior distribution which can avoid over-fitting and lead to a more generalized model.

Unlike LR and SVM, HyBaR incorporates domain knowledge in the form of prior. For the LIWC based features, we have used a sparse prior, and for the category distribution based features we employed the semantic distance between the categories as prior.

The Model: For $i \in [1, \ldots, n]$ we denote $\boldsymbol{s}^i = (s_1^i, \ldots, s_m^i)$ to be the vector that represents the features extracted from the search history of the person p_i, where n is the number of people and m is the number of search categories. For instance, when we consider the search category based features, s_j^i represents the percentage of searches of category j for i^{th} person. Also, let $ls_i \in \{0, 1\}$ be a binary variable denoting whether a person p_i exhibits low self-esteem or not. In particular, $ls_i = 1$ when the person p_i exhibits low self-esteem and 0 otherwise. Now we propose HyBaR, a probabilistic discriminative model with parameters $\boldsymbol{w}(\in \mathbb{R}^m)$ and $\sigma(\in \mathbb{R})$. For each $i \in [1, \ldots, n]$ we assume

$$ls_i \sim \text{Bern}\left(f(\mathbf{w}^T \mathbf{s}^i + \epsilon)\right) \tag{1}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is a latent variable and f is a function whose range is $[0, 1]$ (i.e.which outputs a valid probability). One choice for f can be **Sigmoid**. Also, Bern(p) denotes the bernouli distribution, i.e., $X \sim \text{Bern}(p)$ iff,

$$X = \begin{cases} 1 & \textit{with probability } p \\ 0 & \textit{with probability } (1-p) \end{cases}$$

Next, we describe the prior distribution of the parameters.

Remark: Let $\mathbf{x}$ be the data and y be its label. In case of linear regression, class label, $y(\in \mathbb{R})$ is modeled as,

$$y \sim \mathbf{w}^T \mathbf{x} + \epsilon \tag{2}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and $\mathbf{w}$ is the model parameter. However, for logistic regression, class label $y \in \{0, 1\}$ and is modeled as,

$$y \sim \text{Bern}\left(\textit{Sigmoid}(\mathbf{w}^T \mathbf{x})\right) \tag{3}$$

Hence from (2) and (3) it follows that our model (1) is a simple hybrid of the two models.

Prior Selection: Domain knowledge is incorporated in the model through our priors. Hence, an important component during the model construction is prior selection. The prior of σ is chosen to be uniform in both types of features. However, we set the prior of $\boldsymbol{w}$ depending on the features we are selecting.

For the LIWC based features, the prior is set to be the Laplace distribution, i.e.,

$$P(w_i|\mu, b) = \frac{1}{2b}\exp\left(-\frac{|w_i - \mu|}{b}\right) \tag{4}$$

where μ and b are hyper-parameters. It is to note that Laplace prior is equivalent to sparsity constraint. In the context of LIWC based features, for any individual, these linguistic categories do not co-occur in the search history frequently. This motivated us to select a sparse prior.

For the search category distribution based features, we first construct a similarity matrix that represents the semantic distance between the search categories using GloVe [38]. We set the prior of $\boldsymbol{w}$ to be multivariate Gaussian with mean 0 and the category-similarity matrix as the covariance matrix, i.e.,

$$\boldsymbol{w} \sim \mathcal{N}(\vec{0}, \boldsymbol{Cov}) \tag{5}$$

where $\boldsymbol{Cov}_{ij}$ is the semantic similarity between category i and category j. We set this prior to enforce the similarities between the search categories.

Posterior Inference: Let $S = \left\{\boldsymbol{s}^i | i \in \{1, \ldots, n\}\right\}$ be the set of features for all the participants and $L = \left\{ls^i | i \in \{1, \ldots, n\}\right\}$ be the corresponding self esteem label for them. Now, we formalize the posterior for our model as follows:

$$\begin{aligned} P[\boldsymbol{w}, \sigma^2, \epsilon | L, S] &\propto P[L|\boldsymbol{w}, \sigma^2, \epsilon, S] \cdot P[\boldsymbol{w}, \sigma^2, \epsilon | S] \\ &= \underbrace{P[L|\boldsymbol{w}, \sigma^2, \epsilon, S] \cdot P[\epsilon|\sigma^2]}_{\text{Likelihood}} \cdot \underbrace{P[\boldsymbol{w}] \cdot P[\sigma^2]}_{\text{Prior}} \qquad (6) \\ &= \left(\prod_{i=1}^{n} P[ls_i|\boldsymbol{w}, \sigma^2, \epsilon, \boldsymbol{s}^i]\right) \cdot \left(P[\epsilon|\sigma^2] \cdot P[\boldsymbol{w}] \cdot P[\sigma^2]\right) \qquad (7) \end{aligned}$$

Here (6) follows from conditional independence of S & $\boldsymbol{w}$ and (7) follows from the i.i.d. properties of the data. From (1) we can see that,

$$P[ls_i|\boldsymbol{w}, \sigma^2, \boldsymbol{s}^i] = prob_i^{ls_i} \cdot (1 - prob_i)^{(1-ls_i)}, \tag{8}$$

where, $prob_i = Sigmoid(\mathbf{w}^T \mathbf{s}^i + \epsilon)$. Now, for LIWC based features $P[\boldsymbol{w}]$ in (7) is given by the distribution as in (4) and for the search category based features $P[\boldsymbol{w}]$ in (7) is the distribution as in (5).

In summary, we get the posterior of the parameters from the corresponding prior distribution and the training data likelihood by using Bayes Rule. Then our goal is to estimate the entire posterior distribution and sample model parameters from the posterior distribution. The complete process is depicted in Fig. 5 (b). However, it important to note that exact inference of the posterior given in (7) is intractable. Hence, we use Metropolis-Hastings Algorithm to get approximate sample from the posterior distribution.

Parameter Estimation: Our next step is to choose a parameter for the classification task. We sample several parameters from the posterior distribution as shown in Fig. 5. Let $\{\langle \boldsymbol{w}^1, \sigma^1\rangle, \cdots, \langle \boldsymbol{w}^k, \sigma^k\rangle\}$ be the sampled parameters. Using each of these learned parameters, we classify the test data as follows,

$$ls^i_{test} \sim \text{Bern}\left(Sigmoid(\mathbf{w}^{\mathbf{i}^T} \mathbf{s}_{test} + \epsilon)\right),$$

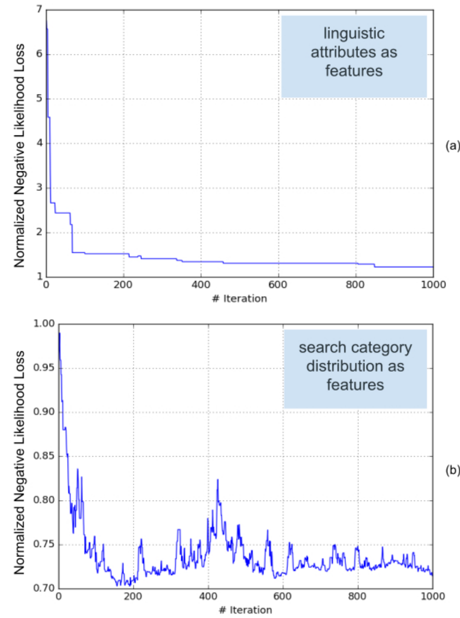

Figure 6: An instance of the convergence of negative log likelihood during training for HyBaR when trained using the two types of features separately.

where $\epsilon \sim \mathcal{N}(0, {\sigma^i}^2)$. Hence we get k labels ($\{ls^1_{test}, \cdots, ls^k_{test}\}$) for each of the data point $\mathbf{s}_{test}$. We record the performance of the model for each $\langle \mathbf{w}^i, \sigma^i\rangle$ and report the average performance of the model on test set in Fig 8 (b) & (d).

Generalization Ability: Following the PAC-Bayesian framework in [17] we can set the normalized negative log likelihood as the loss function for the posterior inference problem. Fig 6 presents an instance of the convergence of negative log likelihood loss during training. Next, we ask the question, how well this model generalizes on a completely new data set? We test the generalization ability of HyBaR by running simulation on large synthetic data. Fig 7 shows the convergence of the expected empirical loss towards the expected true loss. A potential explanation for such behavior can be due to the Bayesian approach used in designing HyBaR.

Driving Search Categories: After training HyBaR on the 27 search categories (from 5.1.2), the top ten categories (ranked according to learned weights) that drive the classification tasks are: Internet & Telecom (2.17), Law & Government (1.91), People & Society (1.85), News (1.73), Real Estate (1.69), Jobs & Education (1.55), Online Communities (1.45), Finance (1.16),Science (1.15), and Health (1.00). High prevalence of these categories in a college student's search history may help identify conditions such as a low self-esteem. For example, searchers pertaining to People & Society categories: how to 'fit in' with communities in college settings, looking for ways to make friends etc. are practical situations that teenagers encounter when they move to college.

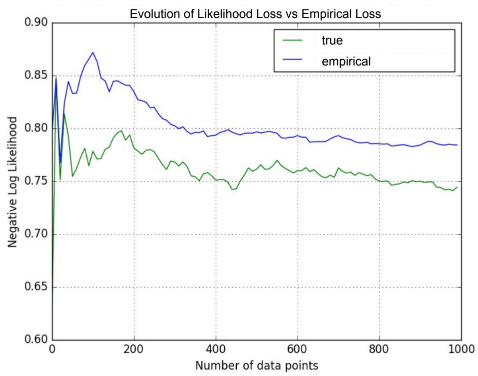

Figure 7: Convergence of the expected empirical loss towards expected true loss on the synthetic data.

6 RESULTS

The performance of our trained models is evaluated using the held out test set which consists of 19 individuals with NLS and 21 with LS condition. Fig. 8 presents the performance of our models. To evaluate the accuracy of the classification, we use F1 score to capture the balance between precision and recall and compare our models.

HyBaR samples multiple parameters from the posterior. We use each of these learned parameters individually to classify our test data and report the average/best performance on the test data (in terms of F1 score).

When LIWC's linguistic attributes are considered as features, on average, HyBaR performed better than both the SVM and the baseline model i.e., LR in classifying participants with low self-esteem. This is evident from Fig. 8 (a) & (b), where HyBaR has an average F1 score of 0.72 compared to SVM's 0.61 and LR's 0.46. We also report the best performance of HyBaR in Fig. 8(b) (numbers with * as superscript) and found that the best model has F1 score of 0.77.

When search category distributions were used as features, HyBaR performed significantly better than SVM and LR, compare Fig. 8 (c) & (d). On average, HyBaR has F1 score of 0.86 compared to SVM's 0.62 and LR's 0.47. Using the category distribution, HyBaR had the best performance with F1 score of 0.89.

Since our test set is small, we compared the performance of the models on the test set by analyzing the confusion matrix. Through Fig. 9, we compare the performance of our models (on the test set) when they are trained using the LIWC based features. The green and the yellow cell in the table represent number of misclassification and HyBaR made the fewest misclassifications compared to LR, our baseline, and SVM. We repeated our analysis on various test set with different ratios of the two categories and obtained similar results.

On average, HyBaR performed significantly better than LR and SVM when we trained it using features based on the search category distribution. The performance of HyBaR on the test set can be evaluated through the confusion matrix presented in Fig. 10. Out of 21 participants with LS condition, HyBaR only misclassified 2 subject (see yellow cell in Fig. 10) and misclassified 4 out of the 19 individuals with NLS condition.

7 DISCUSSION AND FUTURE WORK

Individual level Google search history data can be treated as a rich longitudinal data source that can capture quotidian online activities of individuals. This data has recently become readily accessible through the Google Takeout interface. We built a scalable human-subject review board-approved system for consenting, collecting, and anonymizing subjects' search history data, and created the first data set that links search history data to that from a clinically-used mental health survey instrument.

We showed that search history data provides identifiable signals for detecting low self-esteem. In addition to considering search category distribution, we were able to identify linguistic variables whose frequency of use was significantly different among subjects suffering from low self-esteem and those not exhibiting low self-esteem condition. Using these features, we have managed to build mental-health related predictors, for instance consider our low self-esteem model, that is superior to predicting by random chance. This motivates us to consider that search history data can be used to build models for capturing various mental health phenomena such as self-harm, suicide ideation, etc.

It will be interesting to explore whether both psycho linguistic attributes and search categories can be used simultaneously to build models. Due to small sample size, we could not use both sets of features at the same time. However, this limitation can be easily overcome as we collect more data through our future studies.

The mental health survey we employed red-flagged low self-esteem as the biggest mental health issue for the subject population of college students. Our next study will recruit a more general population and compare the findings. Recruitment sites will include mental-health inpatient and outpatient clinics, emergency rooms, and family court. Individuals from these sites are more likely to have relatively high incidence and prevalence of conditions such as major depression, anxiety, and stress, along with low self-esteem.

We will also consider features beyond linguistic and search categories. Other factors one can consider are time at which these searches are made, seasonality of certain types of searches. The time of a search is likely to be of particular value because searching in the middle of the night can be used as a proxy for sleeplessness, a symptom of many mental health problems. We will test the hypothesis that an increase in a user's volume of searches is a proxy for an increased use of technology due to withdrawal from social interactions.

The fact that search logs are time and date tagged means that they can be used for longitudinal studies on disease progression, as well as part of longitudinal treatment plans for individuals. We are currently beginning a study where subjects who are undergoing outpatient treatment for attempted suicide may consent to allow their history of PHQ-9 survey answers to be analyzed in conjunction with their search histories. The goal of the study will be to model the dynamics of health and online behavior in order to provide new tools for therapists to assess the efficacy of treatment strategies.

Finally, the accessibility of individual level search log data opens the door to studying signals for other mental health phenomena, such as anxiety and panic disorders, bipolar disorder, eating disorders, suicide ideation, and substance abuse and addiction. A feasible use case of our system can be in the university counselling services

Features based on linguistic attributes

	SVM / Logistic Regression (LR)		
	Precision	Recall	F1 Score
NLS	0.45 (0.38)	0.75 (0.50)	0.56 (0.43)
LS	0.70 (0.50)	0.63 (0.48)	0.65 (0.49)
Avg	0.57 (0.44)	0.69 (0.53)	0.61 (0.46)

(a)

	HyBaR		
	Precision	Recall	F1 Score
NLS	<0.57> 0.68*	<0.73> 0.75*	<0.64> 0.71*
LS	<0.92> 0.93*	<0.70> 0.74*	<0.80> 0.82*
Avg	<0.75> 0.81*	<0.72> 0.75*	<0.72> 0.77*

(b)

Features based on search category distribution

	Precision	Recall	F1 Score
NLS	0.58 (0.39)	0.74 (0.50)	0.65 (0.43)
LS	0.69 (0.60)	0.54 (0.45)	0.59 (0.51)
Avg	0.64 (0.49)	0.62 (0.47)	0.62 (0.47)

(c)

	Precision	Recall	F1 Score
NLS	<0.67> 1.0*	<0.66> 0.73*	<0.66> 0.84*
LS	<0.83> 0.90*	<0.94> 1.0*	<0.88> 0.94*
Avg	<0.75> 0.95*	<0.80> 0.87*	<0.86> 0.89*

(d)

*Top number in each cell represent **SVM's** performance.*
*Numbers within ( ) represent **baseline (LR)** model's performance.*

*Numbers within < > represent **HyBaR**'s average performance.*
*Numbers with * represent **HyBaR** model's best performance.*

Figure 8: Classifiers performance on the Test Set. LR, SVM, and HyBaR are trained for the classification task. (a) & (b): Tables present the performance on the test set when LIWC based features are used. (c) & (d): Tables present the performance on the test set when search category based features are used. In case of HyBaR, the model learns multiple parameters (see From Fig 5), the test set is evaluated using each of the parameters. Numbers in ⟨⟩ represents the average performance over the learned parameters. Numbers with * represent the score of the best performing HyBaR parameter on the test set.

N = 40 (NLS: 19, LS: 21)	Predicted NLS	Predicted LS
True NLS	LR: 5 SVM: 14 HyBaR: 15	LR: 14 SVM: 5 **HyBaR: 4**
True LS	LR: 12 SVM: 8 **HyBaR: 5**	LR: 9 SVM: 13 HyBaR: 16

Figure 9: Confusion matrix to compare the performance of Logistic Regression(LR), Support Vector Machine (SVM), and HyBaR (average performance of HyBaR) on the test set when trained using linguistic features.

N = 40 (NLS: 19, LS: 21)	Predicted NLS	Predicted LS
True NLS	LR: 7 SVM: 14 HyBaR: 15	LR: 12 SVM: 5 **HyBaR: 4**
True LS	LR: 6 SVM: 10 **HyBaR: 2**	LR: 15 SVM: 11 HyBaR: 19

Figure 10: Confusion matrix to evaluate the performance of LR, SVM, and HyBaR on test set when trained on features based on search category distribution.

to access the morale of the student population which can eventually lead to better understanding of student life.

We are working with other researchers and clinicians at our university's medical center to include this work as a core component of a suicide prevention center that is currently being developed. For example, this is now a central feature of the proposed ALACRITY Research Center focused on suicide prevention among patients with Serious Mental Illness at our institution's medical center[4]. It will form a core data component for use in all NIH-funded sub-projects within the Center. This offers to expand the overall social impact by influencing clinical care in the future for those individuals who are among the absolute highest risk for suicide mortality.

ACKNOWLEDGMENTS

We thank Adam Sadilek and other collaborators at Google and Verily for discussions, advice, and the use of anonymization software. This work was supported by the Goergen Institute for Data Science and Google.

REFERENCES

[1] Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2015. Depression-related imagery on instagram. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative work & social computing*. ACM, 231–234.

[2] John W Ayers, Benjamin M Althouse, Jon-Patrick Allem, J Niels Rosenquist, and Daniel E Ford. 2013. Seasonality in seeking mental health information on Google. *American journal of preventive medicine* 44, 5 (2013), 520–525.

[4]https://grants.nih.gov/grants/guide/pa-files/PAR-16-354.html

[3] John W Ayers, Benjamin M Althouse, Eric C Leas, Mark Dredze, and Jon-Patrick Allem. 2017. Internet searches for suicide following the release of 13 Reasons Why. *JAMA internal medicine* 177, 10 (2017), 1527–1529.

[4] Roy F Baumeister. 1990. Suicide as escape from self. *Psychological review* 97, 1 (1990), 90.

[5] Sunil Bhar, Marjan Ghahramanlou-Holloway, Gregory Brown, and Aaron T Beck. 2008. Self-esteem and suicide ideation in psychiatric outpatients. *Suicide and Life-Threatening Behavior* 38, 5 (2008), 511–516.

[6] Jiang Bian, Umit Topaloglu, and Fan Yu. 2012. Towards large-scale twitter mining for drug-related adverse events. In *Proceedings of the 2012 international workshop on Smart health and wellbeing*. ACM, 25–32.

[7] Christopher M. Bishop. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

[8] Ronald L Bonner and Alexander Rich. 1988. Negative life stress, social problem-solving self-appraisal, and hopelessness: Implications for suicide research. *Cognitive Therapy and Research* 12, 6 (1988), 549–556.

[9] Herman Anthony Carneiro and Eleftherios Mylonakis. 2009. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical infectious diseases* 49, 10 (2009), 1557–1564.

[10] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3267–3276.

[11] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference*. ACM, 47–56.

[12] Munmun De Choudhury, Scott Counts, Eric J Horvitz, and Aaron Hoff. 2014. Characterizing and predicting postpartum depression from shared facebook data. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 626–638.

[13] Munmun De Choudhury and Sushovan De. 2014. Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity.. In *ICWSM*.

[14] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting Depression via Social Media.. In *ICWSM*. 2.

[15] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, 2098–2110.

[16] Johannes C Eichstaedt, Robert J Smith, Raina M Merchant, Lyle H Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A Asch, and H Andrew Schwartz. 2018. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences* (2018), 201802331.

[17] Pascal Germain, Francis Bach, Alexandre Lacoste, and Simon Lacoste-Julien. 2016. Pac-bayesian theory meets bayesian inference. In *Advances in Neural Information Processing Systems*. 1884–1892.

[18] Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. 2009. Detecting influenza epidemics using search engine query data. *Nature* 457, 7232 (2009), 1012.

[19] Google. 2017. Google Trends. https://trends.google.com/trends/

[20] John F Gunn III and David Lester. 2013. Using google searches on the internet to monitor suicidal behavior. *Journal of affective disorders* 148, 2 (2013), 411–412.

[21] Akihito Hagihara, Shogo Miyazaki, and Takeru Abe. 2012. Internet suicide searches and the incidence of suicide in young people in Japan. *European archives of psychiatry and clinical neuroscience* 262, 1 (2012), 39–46.

[22] Jeffrey A. Hall and Natalie Pennington. 2013. Self-monitoring, honesty, and cue use on Facebook: The relationship with user extraversion and conscientiousness. *Computers in Human Behavior* 29, 4 (2013), 1556 – 1564.

[23] Carl L Hanson, Scott H Burton, Christophe Giraud-Carrier, Josh H West, Michael D Barnes, and Bret Hansen. 2013. Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. *Journal of medical Internet research* 15, 4 (2013).

[24] Catherine Hoffman and Julia Paradise. 2008. Health insurance and access to health care in the United States. *Annals of the New York Academy of Sciences* 1136, 1 (2008), 149–160.

[25] Becky Inkster, David Stillwell, Michal Kosinski, and Peter Jones. 2016. A decade into Facebook: where is psychiatry in the digital age? *The Lancet Psychiatry* 3, 11 (2016), 1087–1090.

[26] Sue Jamison-Powell, Conor Linehan, Laura Daley, Andrew Garbett, and Shaun Lawson. 2012. I can't get no sleep: discussing# insomnia on twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1501–1510.

[27] Lauren A Jelenchick, Jens C Eickhoff, and Megan A Moreno. 2013. âĂIJFacebook depression?âĂİ Social networking site use and depression in older adolescents. *Journal of Adolescent Health* 52, 1 (2013), 128–130.

[28] John Kolligian Jr and Robert J Sternberg. 1991. Perceived Fraudulence in Young Adults: Is There an'Imposter Syndrome'? *Journal of personality assessment* 56, 2 (1991), 308–326.

[29] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. 2001. The phq-9. *Journal of general internal medicine* 16, 9 (2001), 606–613.

[30] Mingming Liu, Jia Xue, Nan Zhao, Xuefei Wang, Dongdong Jiao, and Tingshao Zhu. 2018. Using social media to explore the consequences of domestic violence on mental health. *Journal of interpersonal violence* (2018), 0886260518757756.

[31] Michael J McCarthy. 2010. Internet monitoring of suicide risk in the population. *Journal of affective disorders* 122, 3 (2010), 277–279.

[32] Blace A Nalavany and Lena W Carawan. 2012. Perceived family support and self-esteem: The mediational role of emotional experience in Adults with Dyslexia. *Dyslexia* 18, 1 (2012), 58–74.

[33] Sudhakar V Nuti, Brian Wayda, Isuru Ranasinghe, Sisi Wang, Rachel P Dreyer, Serene I Chen, and Karthik Murugiah. 2014. The use of google trends in health care research: a systematic review. *PloS one* 9, 10 (2014), e109583.

[34] David Nutt, Sue Wilson, and Louise Paterson. 2008. Sleep disorders as core symptoms of depression. *Dialogues in clinical neuroscience* 10, 3 (2008), 329.

[35] Bridianne O'dea, Mark E Larsen, Philip J Batterham, Alison L Calear, and Helen Christensen. 2017. A linguistic analysis of suicide-related Twitter posts. *Crisis* (2017).

[36] Andrew Page, Shu-Sen Chang, and David Gunnell. 2011. Surveillance of Australian suicidal behaviour using the Internet?

[37] JW Pennebaker, RJ Booth, RL Boyd, and ME Francis. 2015. Linguistic inquiry and word count: LIWC2015. Austin, TX: Pennebaker Conglomerates.

[38] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[39] Carly R Pineiro. 2016. *Social media use and self-esteem in undergraduate students*. Rowan University.

[40] Wullianallur Raghupathi and Viju Raghupathi. 2014. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems* 2, 1 (2014), 3.

[41] Richard W Robins, Holly M Hendin, and Kali H Trzesniewski. 2001. Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and social psychology bulletin* 27, 2 (2001), 151–161.

[42] Rebecca L Robinson, Reanelle Navea, and William Ickes. 2013. Predicting final course performance from studentsâĂŹ written self-introductions: A LIWC analysis. *Journal of Language and Social Psychology* 32, 4 (2013), 469–479.

[43] Morris Rosenberg. 1965. Rosenberg self-esteem scale (RSE). *Acceptance and commitment therapy. Measures package* 61 (1965), 52.

[44] Amira Y Sharaf, Elaine A Thompson, and Elaine Walsh. 2009. Protective effects of self-esteem and family support on suicide risk behaviors among at-risk adolescents. *Journal of Child and Adolescent Psychiatric Nursing* 22, 3 (2009), 160–168.

[45] Morton M Silverman and Ronald W Maris. 1995. The prevention of suicidal behaviors: An overview. *Suicide and Life-Threatening Behavior* 25, 1 (1995), 10–21.

[46] Aaron Smith and Monica Anderson. 2018. Social Media Use in 2018. http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/. Accessed 7 June 2018.

[47] Tae Min Song, Juyoung Song, Ji-Young An, Laura L Hayman, and Jong-Min Woo. 2014. Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends. *Yonsei medical journal* 55, 1 (2014), 254–263.

[48] Julia Friederike Sowislo and Ulrich Orth. 2013. Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological Bulletin* 139, 1 (Jan. 2013), 213–240.

[49] Hajime Sueki. 2011. Does the volume of Internet searches using suicide-related search terms influence the suicide death rate: Data from 2004 to 2009 in Japan. *Psychiatry and clinical neurosciences* 65, 4 (2011), 392–394.

[50] Michael Van Ameringen, Catherine Mancini, William Simpson, and Beth Patterson. 2010. Potential use of Internet-based screening for anxiety disorders: a pilot study. *Depression and anxiety* 27, 11 (2010), 1006–1010.

[51] Victor R. Wilburn and Delores E. Smith. 2005. STRESS, SELF-ESTEEM, AND SUICIDAL IDEATION IN LATE ADOLESCENTS. *Adolescence* 40, 157 (Spring 2005), 33–45. https://search.proquest.com/docview/195943093?accountid=13567 Copyright - Copyright Libra Publishers Incorporated Spring 2005; Document feature - Tables; ; Last updated - 2017-10-31; CODEN - ADOLAO.

[52] Albert C Yang, Norden E Huang, Chung-Kang Peng, and Shih-Jen Tsai. 2010. Do seasons have an influence on the incidence of depression? The use of an internet search engine query data as a proxy of human affect. *PloS one* 5, 10 (2010), e13728.

[53] Albert C Yang, Shi-Jen Tsai, Norden E Huang, and Chung-Kang Peng. 2011. Association of Internet search trends with suicide death in Taipei City, Taiwan, 2004–2009. *Journal of affective disorders* 132, 1 (2011), 179–184.

[54] Shan Yin and Mona Ho. 2012. Monitoring a toxicological outbreak using Internet search query data. *Clinical toxicology* 50, 9 (2012), 818–822.

[55] Elad Yom-Tov and Shira H Fischer. 2017. The Werther Effect Revisited: Measuring the Effect of News Items on User Behavior. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 1561–1566.
[56] Soo Jeong Youn, Nhi-Ha Trinh, Irene Shyu, Trina Chang, Maurizio Fava, Joseph Kvedar, and Albert Yeung. 2013. Using online social media, Facebook, in screening for major depressive disorder among college students. *International Journal of Clinical and Health Psychology* 13, 1 (2013), 74–80.