Title: A Logistic Regression Classifier for Suicidal Reddit Posts

Introduction:

Though suicide is one of the leading causes of death in many age groups (CITATION), death by suicide is deeply stigmatized and finding adequate support is often difficult for suicidal individuals (CITATION). However, studies have shown that people are more willing to share thoughts and feelings online than in person (Lenhart et al., 2001; McKenna & Bargh, 2000), making the internet a powerful resource for individuals to frankly discuss thoughts of suicide. In this project, we examine the expression of suicidal ideation on Reddit, a popular online message board aggregator that hosts discussion on user-created topics in smaller sub-boards (known as "subreddits"). Notably, some subreddits are used to provide and ask for mental health support, and many users share deeply personal information, including thoughts of suicide.

Here, we develop a logistic regression model that scores text submissions and comments from mental health subreddits for the presence of suicidal ideation. Previous work on suicidal ideation among Reddit users incorrectly assumed that popular mental health subreddits other than r/SuicideWatch, a subreddit dedicated to expressions of intent to commit suicide, do not contain suicidal posts (de Choudhury 2015). In reality, the proportion of suicidal posts in r/depression, the most popular mental health subreddit, is at least 8%. **As such, developing reliable identification methods for suicidal ideation in text posts is an important step to improve future work on suicidal ideation on internet discussion boards**. Our analysis suggests that…[to be completed]

Methods:

Data Collection

The analysis was performed on text submissions and posts obtained from several mental health subreddits. Data were extracted from Reddit JSON files using a JSON parser (written by Lacey and John) that takes in a JSON file and outputs a data frame that contains the attributes of a submission, along with all of its replies (hereafter referred to as comments).

First, we obtained ~110 submissions (i.e. comments from the original poster), randomly sampled from a set of ~20 of the largest mental health subreddits excluding r/depression, using the PRAW python package's random sampling function (scraped by Lacey). Comments were scored (by both Lacey and Linda) for current suicidal ideation, past suicidal ideation, and the presence/absence of mood or anxiety disorders. ~50% of these posts indicate suicidal ideation in some point in time.

Secondly, a large training set of ~400 comments from mental health subreddits, including but not limited to r/depression. The comments were obtained and scored as follows. ~400 comments (consisting of both submissions and replies) from r/depression that were posted in June (scraped by Linda). Lacey and Linda each scored ~300 comments, with ~100 comments scored by both. Comments were scored only for the presence of suicidal ideation, and were obtained from June posts in r/depression. Only about 30 of these comments indicate suicide ideation.

Using these initial training sets, we (Lacey and Linda) developed a lexicon of common phrases that indicate suicidal ideation or intention, created based on the phrases that indicated suicidal ideation in the suicidal posts (scored by Lacey and Linda). We used this to write an initial classification algorithm (written by Lacey and Linda) that searches individual posts for phrases that match the lexicon phrases that indicate suicidal ideation, which were identified from the suicidal posts that were hand-classified. (The algorithm was written to have very high sensitivity (GIVE EXACT NUMBER) in order to enrich the set of suicidal users, but the specificity of the algorithm was questionable.)

Thirdly, all the comments from r/depression from June 1-30, 2016, in JSON format (scraped by Linda) and were scored for suicidal ideation by the phrase-matching algorithm to enrich the set of posts and users with suicidal ideation. The algorithm performed well, identifying ~300 suicidal submissions in a set of ~5000 submissions, with only one false positive.

Finally, to create the dataset for building the model, we combined:

-the hand-scored submissions from multiple mental health subreddits

-the hand-scored submissions and comments from r/depression in June, for which

The suicidal scores matched for both Lacey and Linda

-the algorithm-identified suicidal submissions from r/depression in June

resulting in a training set of ~600 comments, ~400 suicidal and ~200 non-suicidal.
(may add in algorithm-identified suicidal comments later)

Exploratory Analysis

Exploratory analysis was performed by examining tables and plots of the observed data. Exploratory data analysis was used to perform quality control on the data and determine the predictor variables for the the logistic regression model. Specifically, we sought to confirm the existence of the characteristics identified by de Choudhury et al (2015) as possibly indicating suicidal ideation in text posts, which we elaborate on in the results section.

Statistical Modeling

Here, we use logistic regression as the analysis framework. Although there are several other binary classification algorithms, such as support vector machines, discriminant analysis, and random forests, each of which suffers drawbacks that make them unsuitable to classify this dataset. Support vector machines aren't well suited to this problem, because it works better when the predictors clearly separate the classes (CITATION), which isn't true here because many of the predictor values are similar between suicidal and non-suicidal texts except for outliers. Discriminant analysis performs better when the number of observations in both groups is relatively equal (CITATION), whereas the number of suicidal posts (663) exceeded the number of the number of non-suicidal posts (400) by a substantial margin in our training set. Random forests, while potentially useful for determining more complex relationships between predictors and the classification, are difficult to interpret (CITATION), and thus the insights gleaned from such a model may be of limited public health use. As such, logistic regression was

ultimately chosen as it is a relatively easy-to-understand framework and suffers no serious drawbacks due to the data available.

The predictive variables were chosen on the basis of the exploratory analysis, prior knowledge of characteristics of suicidal posts, and the drop-in-deviance test. Coefficients were estimated by . Standard errors were calculated by.

Reproducibility
To be continued.

Results:
Previous work by de Choudhury (2015) identified several characteristics of the posts that users who eventually post in SuicideWatch. Here, we first seek to confirm that the same patterns are upheld in users that make suicidal posts and comments to mental health subreddits that aren't SuicideWatch. Secondly, we use the promising variables create multiple logistic regression models, and calculate the drop-in-deviance for each to choose the best model. **To test which model , we use k-fold cross validation to determine which model is most generalizable**. Finally, we compare the sensitivity and specificity of the best logistic regression model to the initial classification algorithm, based on a fresh random sample of posts from r/depression (scraped by Linda).

Linguistic Structure
Posts from suicidal users have decreased readability, possibly indicating decreased cognitive function and coherence (de Choudhury (2015,). This pattern is confirmed in our data: (Fig.). Furthermore, suicidal users use more verbs and adverbs, and fewer nouns. (not analyzed yet, might skip).

Interpersonal Differences
Suicidal users' posts have more first-person pronouns, and fewer second person and third person pronouns. This may indicate greater self-attentional focus, a known risk factor for depression and suicide (de Choudhury 2015,). This pattern is (not?) confirmed in our data: (started analysis, not yet complete).

Interaction Differences
Suicidal users have longer posts, which indicate increased self-disclosure (de Choudhury 2015,). This pattern is confirmed in our data: (Fig.).

Content
Word Choice - Some interesting differences, see figure.

Sentiment - Sentiment was not used as a predictor in de Choudhury's work, but previous work on Twitter has shown that sentiment is useful predictor for depressed/suicidal users

(CITATION). We use two packages: the nrc sentiment library from; and RSentiment, an R package that simply classifies text as being positive, negative, neutral, or sarcastic.

Predictably, the sentiment of suicidal users was much more negative (Fig.) . Interestingly, several suicidal posts (as well as non-suicidal posts) were marked as being sarcastic.

Model selection and comparison

Discussion and Conclusion:
Our analysis suggests that .

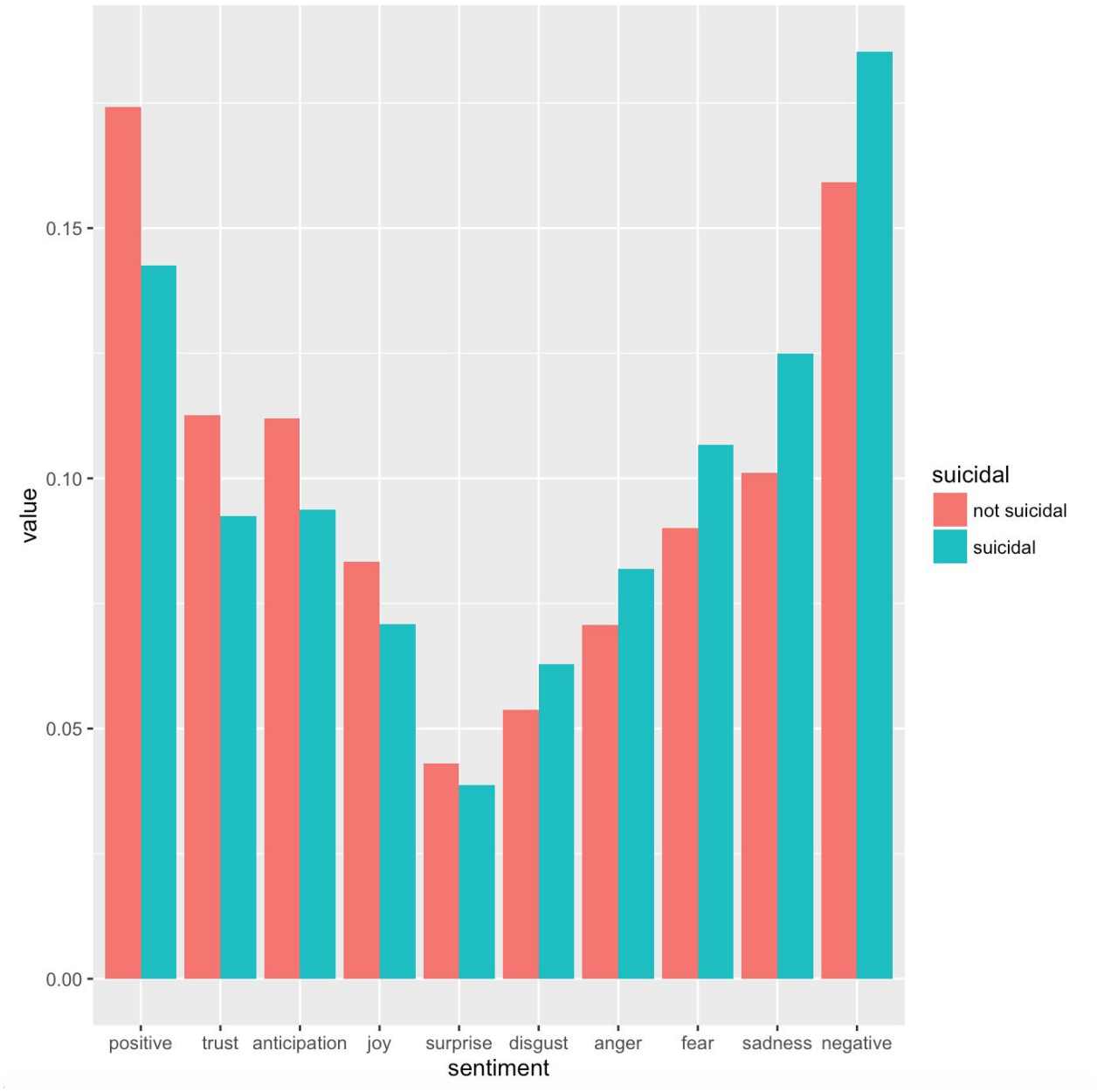
Our analysis is limited in that .
Future work:
Here, we used only the text of the post to predict suicidal ideation. Other possible predictors:

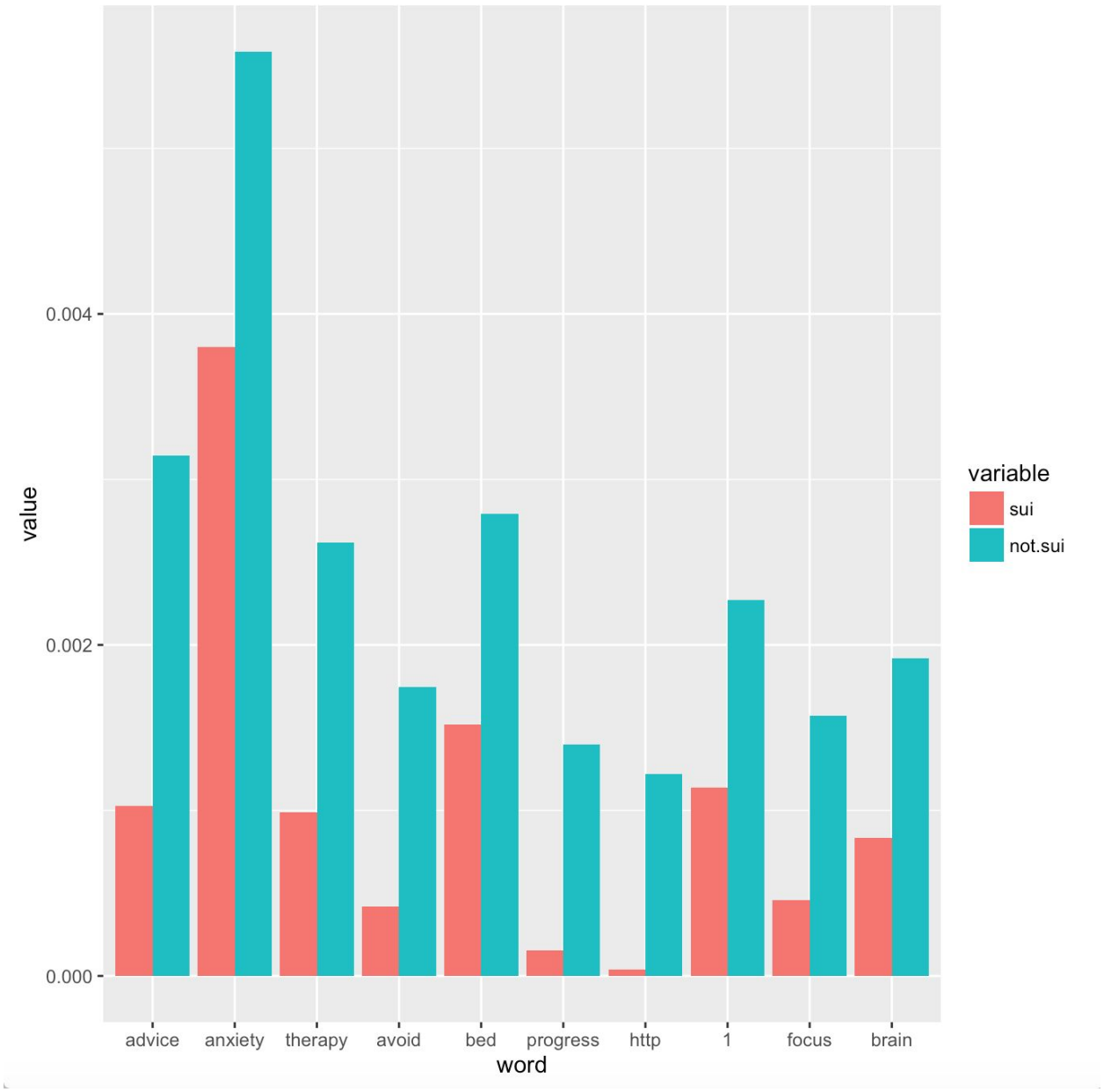
-Number of comments
-Pattern differences between throwaway and regular accounts

Develop an algorithm that predicts whether a user will make a suicidal post, based on their post history and past behavior

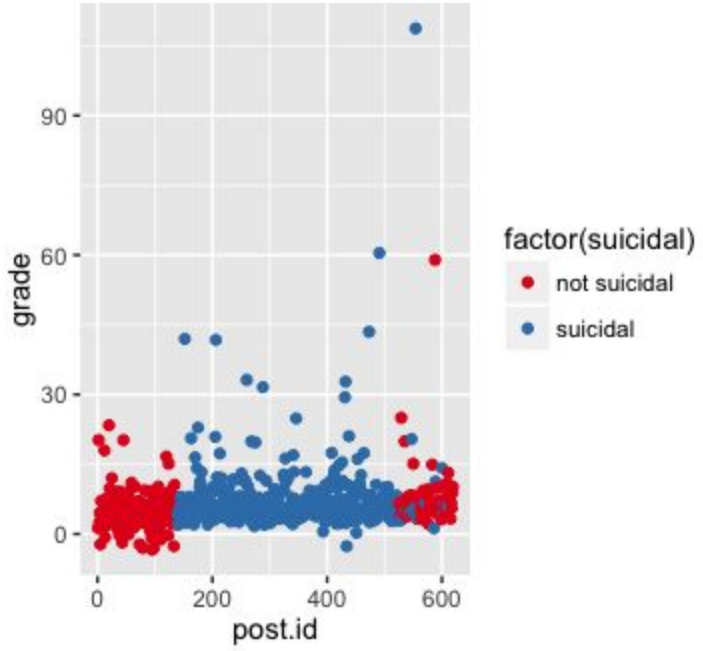
However, our conclusion is valuable because . While a recent paper characterized users that would go on to express intent to commit suicide versus those that remained in the mental health subreddits (CITATION), its findings were limited in that the authors assumed that posters in mental health subreddits that weren't SuicideWatch did not express suicidal thoughts. However, a preliminary algorithm identified that at least 8% of the posts in r/depression contain thoughts of suicide. The field of internet-based mental health research is young and is a rich area of future research and opportunities to improve support for people with mental illnesses.

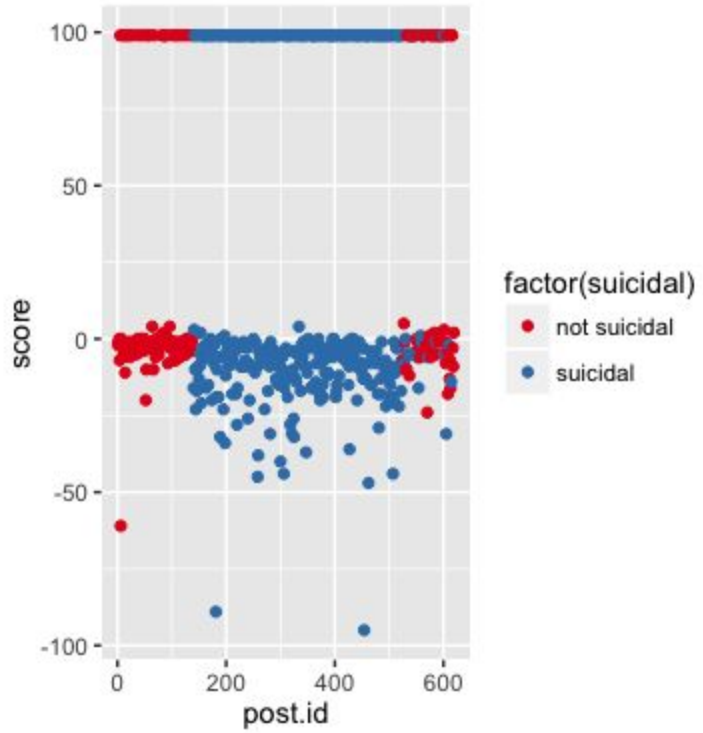
Figures

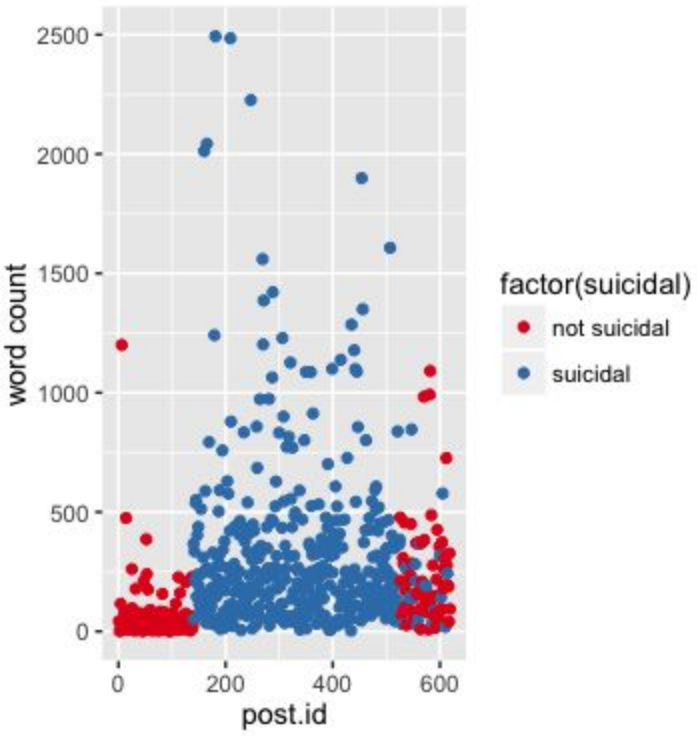

0.15
0.10
0.05
0.00
value
positive
trust
anticipation
joy
surprise
disgust
anger
fear
sadness
negative
sentiment
suicidal
not suicidal
suicidal


0.004
0.002
0.000
value
advice
anxiety
therapy
avoid
bed
progress
http
1
focus
brain
word
variable
sui
not.sui

lesch-Kincaid Readability of posts and
comments in mental health subreddits






2500
2000
1500
1000
500
0
word count
0
200
400
600
post.id
factor(suicidal)
not suicidal
suicidal