

Detecting anxiety on Reddit

Judy Hanwen Shen

University of Toronto
Toronto, Canada

judyhanwen.shen@mail.utoronto.ca

Frank Rudzicz

Toronto Rehabilitation Institute-UHN
University of Toronto

Toronto, Canada
frank@spoclab.com

Abstract

Previous investigations into detecting mental illnesses through social media have predominately focused on detecting depression through Twitter corpora (De Choudhury et al., 2013; Resnik et al., 2015; Pedersen, 2015). In this paper, we study anxiety disorders through personal narratives collected through the popular social media website, Reddit. We build a substantial data set of typical and anxiety-related posts, and we apply N -gram language modeling, vector embeddings, topic analysis, and emotional norms to generate features that accurately classify posts related to binary levels of anxiety. We achieve an accuracy of 91% with vector-space word embeddings, and an accuracy of 98% when combined with lexicon-based features.

1 Introduction

Anxiety disorders include a family of conditions characterized by excessive fear, emotional responses to real or perceived threats, and worry in anticipation of future threats. Common forms of anxiety include generalized anxiety, social anxiety, health anxiety, and panic attacks (American Psychiatric Association, 2013). The World Health Organization estimates the 12-month prevalence of anxiety disorders to be 26.4% in the United States (Demyttenaere et al., 2004). In adolescents aged 13-18, anxiety disorders are the most common condition with a lifetime prevalence of 31.9% for all anxiety disorders and 8.9% for severe anxiety disorders (Merikangas et al., 2010).

Anxiety disorders are primarily diagnosed by physicians or psychologists, but 77% of counties in the United States have a severe shortage

of psychiatrists and non-prescribing mental health providers such as psychiatric nurses, social workers, licensed professionals, counselors, and marriage and family therapists (Thomas et al., 2009). Given the high prevalence of these disorders, and the shortage of relevant mental health professionals, there is an urgent need for mental health detection tools that are scalable to large populations, and that can be made widely accessible. In particular, the high prevalence of anxiety disorders in adolescents motivates building these screening tools on emerging social media and communication platforms.

2 Background

Social media has become an increasingly popular data source for detecting mental illnesses through text. For example, De Choudhury et al. (2013) built a corpus of more than 2 million Twitter posts, including a ‘depression’ class with tweets from 476 highly active users self-identified as clinically diagnosed with depression. To identify depression, they used feature vectors that included engagement with the Twitter platform, the social graph of user Twitter activity, emotional and linguistic style using Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2015) and a depression lexicon including antidepressant names. De Choudhury et al. (2013) used a mix of text features and metadata features and achieved 70% accuracy in predicting depression in tweets. Another major data set of tweets labelled for depression was generated by Coppersmith et al. (2015), and contained 3 million tweets from about 2000 Twitter users, including 600 self-identified clinically depressed users. From this data set, Nadeem (2016) achieved 86% accuracy with a naïve Bayes unigram classifier. Resnik et al. (2015) used the same data set with latent Dirichlet allocation

(LDA) and supervised LDA techniques to predict the likelihood of target classes based on topics. Their supervised LDA techniques included the associated labels of documents as priors for topic modeling. This approach modified an unsupervised learning method and achieved a precision of 0.648 at a recall of 0.5. [Preotiuc-Pietro et al. \(2015\)](#) also participated in the shared task and applied a range of methods including: LDA, word vector embeddings, GloVe vector embeddings, and unigrams in order to generate word clusters and then feature vectors based on said word clusters. The same data set has also been used to identify patients with post-traumatic stress disorder (PTSD) in social media in the Coppersmith shared task ([Coppersmith et al., 2015](#)). Using this data set, [Pedersen \(2015\)](#) used lexical decision lists with N -grams (N between 1 and 6) and achieved a classification accuracy of 74.2% in classifying tweets from people with PTSD.

While Twitter data are available in large volumes, tweets are limited in length and can restrict the potential for contextual processing. By contrast, LiveJournal is a platform for people to discuss common interests, and has also been studied to identify community posts by people with depression. [Nguyen et al. \(2014\)](#) found that affective word features from the Affective Norms for English Words (ANEW) and mood tags posted by users gave lower coverage than LIWC features and LDA Topic modeling. Using LIWC and LDA as features for classification, they achieved 93% accuracy.

Psychopathology researchers have investigated social anxiety in the context of social media. For example, [Fernandez et al. \(2012\)](#) studied profile information and usage patterns of Facebook users. They concluded that social anxiety was significantly negatively correlated with the number of Facebook friends and positively correlated with the number of completed sections of a Facebook profile.

Similar to LiveJournal and Facebook, Reddit offers relatively rich bodies of text from users in the context of self-assembled communities. Reddit is a social website for news aggregation, content rating, and discussion. Reddit allows posts up to 40,000 characters per comment, compared to the 140-character limit of Twitter. Each month, 234 million unique users contribute 75.15 million

posts and 725.85 comments to the site ¹. The website contains more than 1 million subpages, called subreddits, each focusing on its own topic, many of which involve sharing personal stories and experiences in order to seek or give advice. The subreddits concerning depression and anxiety both involve over 100,000 community members ². [De Choudhury and De \(2014\)](#) studied mental health disclosure on Reddit and concluded that users share their experiences and challenges with mental illnesses as well as the impacts of their illnesses on their work, lives, and relationships. They also found that users use the platform not only for self-expression, but also for seeking diagnosis and treatment information for their conditions. [Kumar et al. \(2015\)](#) studied the `r/SuicideWatch` community on Reddit after celebrity suicides and found increased posting activity and increased suicidal ideation in post content, by using linguistic measures, N -gram comparison, and topic modeling.

Previous success in detecting depression on social media, combined with the previous qualitative research in anxiety on social media, suggest that there is potential for detecting anxiety and anxious behavior on social media. In this paper, we make the first attempt to detect anxiety-related posts from Reddit using various linguistic features. Specifically, we investigate the effectiveness of vector-space representations and LDA features, compared to LIWC and N -gram models, in distinguishing anxiety-related texts from more typical texts.

3 Data Collection

The extensive Reddit API allows direct access to posts by subreddit. For this experiment, we collected 22,808 posts on Reddit over 3 months. These posts include 9971 anxiety-related posts ('Anxiety') and 12,837 general posts ('Control'). The Anxiety posts are predominantly collected from `r/anxiety`; three other anxiety-related subreddits, including `r/panicparty`, `r/healthanxiety`, and `r/socialanxiety`, are also mined for posts for the Anxiety class. Since the anxiety-related posts are overwhelmingly from a first-person point-of-view, we also collected posts for the Control class from a variety of different subreddits

¹about.reddit.com (2017)

²reddit.com/r/depression, reddit.com/r/anxiety (2017)

| anxiety subreddits | control subreddits | |
|--------------------|---------------------|------------------------|
| r/anxiety | r/askscience | r/relationships |
| r/healthanxiety | r/writingprompts | r/teaching |
| r/socialanxiety | r/writing | r/parenting |
| r/panicparty | r/atheism | r/christianity |
| | r/showerthoughts | r/jokes |
| | r/lifeprotips | r/writing |
| | r/personalfinance | r/talesfromretail |
| | r/theoryoffreddit | r/talesfromtechsupport |
| | r/randomkindness | r/talesfromcallcenters |
| | r/books | r/fitness |
| | r/askdocs | r/frugal |
| | r/legaladvice | r/youshouldknow |
| | r/nostupidquestions | |

Table 1: Subreddits used for data collection.

that involve first-person narratives. Using a diverse mix of subreddits also minimizes the impact of subject-specific words from any given community. Table 1 lists the subreddits included in each category of data. In the Anxiety collection, the average length of posts was 171.83 words (869.14 characters). In the Control posts group, the average length was 164.82 words (846.28 characters). These counts reflect the number of processed tokens, with URLs, HTML tags and punctuation removed. We apply further preprocessing by removing stop words and lemmatizing word tokens.

4 Feature Generation

4.1 Vector space embeddings: Word2Vec and Doc2Vec

Mikolov et al. (2013) introduced an efficient estimation of words in vector space for both skip-gram and continuous bag-of-words (CBOW) models. With all training examples, we constructed a CBOW model with a window size of 5 words between current and predicted words in the sentence, and use the mean of the context word vectors. For training, we make 5 iterations over the corpus and use negative down sampling to draw 5 noise words to speed up training. We empirically select an embedding dimension of 300. With the CBOW model, we constructed feature vectors by taking the mean of all tokens in each training example. Intuitively, this corresponds to finding the center of the cluster of words in the vector space belonging to the target label category.

Predictive models can be further strengthened by incorporating paragraph context. Le and Mikolov (2014) introduced a distributed memory model with paragraph vectors (PV-DM). Each paragraph vector was mapped to a unique vector in addition to each word being mapped to a

unique vector. In the present work, during training of the feature-generation model, in addition to word vector updates, paragraph vectors are inferred with each new training example using gradient descent. The paragraph vectors are used in addition to the word vectors to build the post’s feature vector. Fixed length contexts are computed using a sliding window over the paragraph. The contexts produce paragraph information which act as a memory component to provide history when predicting the next word. We construct a PV-DM model with a window size of 10 and again empirically select an embedding dimension of 300 for all training example, and use negative down sampling to draw 5 noise words. To increase model representation capacity, we iterate over the corpus 10 times. We use the average of the paragraph and word vectors for classification. After generating the PV-DM model, we infer the feature vector using the model by averaging the paragraph vector with the vector representations with the other words in the sentence for each training example.

4.2 LDA topic modeling

| Topic | Topic Words | IG |
|----------|-----------------------------------|--------|
| Ctrl. T1 | like, want, know, go, said | 0.1764 |
| Anxi. T4 | try, drive, look, car, walk | 0.0850 |
| Anxi. T2 | drink, smoke, alcohol, weed, draw | 0.0792 |
| Ctrl. T2 | pay, money, will, can, account | 0.0699 |
| Ctrl. T8 | year, will, time, work, school | 0.0691 |
| Anxi. T3 | school, class, year, college | 0.0654 |
| Anxi. T8 | game, help, eat, food, play | 0.0654 |
| Anxi. T5 | work, job, get, call, time | 0.0654 |
| Anxi. T0 | people, feel, social, think | 0.0654 |
| Ctrl. T1 | doctor, pain, medic, feel, feel | 0.0654 |

Table 2: Topics from the LDA model with the highest information gain (IG) in bits.

Latent Dirichlet allocation (LDA) is a Bayesian generative technique that models bodies of text as

a mixture of underlying latent topics where each topic is characterized by a distribution over individual words (Blei et al., 2003). First, we use the training set to generate two LDA models for the Control and Anxiety classes, respectively. After training the LDA model, we generate the latent unlabelled topics for each class. Table 2 shows the 10 topics, across both groups, with the highest information gain.

Here, each training example is represented by a 20-dimensional array of likelihoods generated by the top 10 topics for each of the Anxiety and Control LDA models.

4.3 LIWC features

| Attributes | IG |
|-------------------|--------|
| Anxiety | 0.3939 |
| Negative | 0.2260 |
| Dictionary words | 0.1815 |
| Affect | 0.1396 |
| First person sing | 0.1373 |
| Emotional tone | 0.1158 |
| Authentic | 0.1112 |
| Clout | 0.0988 |
| Analytic | 0.0783 |
| Feel | 0.0766 |

Table 3: Features from LIWC 2015 with highest information gain (IG) in bits.

We use LIWC 2015 (Pennebaker et al., 2015) to extract lexico-syntactic features as a baseline measure, and the default LIWC dictionary with 95 categories to generate the feature vectors. Table 3 shows the top 10 features from the 94 features of LIWC 2015 with the highest information gain in our data.

4.4 N -gram language models

Another standard method used to extract features from text is to calculate the probability of a document within a language model. In our experiments, we use four different corpora to calculate probabilities of unigrams and bigrams. We build the first two models using the Anxiety and Control training examples, respectively. We build the third model using 100,000 unlabelled tweets from the Sentiment140 dataset (Go et al., 2009) and use the NLTK Brown corpus (Bird, 2006) for the fourth model. To generate feature vectors, we calculate the log-probability of each input sentence as unigrams and as bigrams. For each Reddit post, the associated feature subvector contains a unigram

and bigram probability, with Laplace smoothing for each model, i.e., 8 dimensions in total.

4.5 Learning embeddings and topics

The type of model from which we extract features can be built using any corpus. Here, we compare using in-domain training examples with using another corpus for building the word vector (with word2vec), document vector (with doc2vec), and topic (LDA) models. Here, we choose Twitter as a suitable candidate, since it constitutes a similar social media platform, and since previous literature used Twitter data. To compare, 100,000 tweets from Sentiment140 (Go et al., 2009) were used to build word2vec, doc2vec and LDA topic models. These models were further used to generate training and test feature vectors. Table 4 summarizes the different accuracies of using our Reddit training set compared to using Twitter data to build the feature generation models. Higher accuracies were achieved when the models were trained with Reddit examples rather than with the 100,000 tweets for word2vec and LDA features. However, the Twitter-trained document vector model generated more effective feature vectors than the equivalent model from Reddit data. This result is likely due to the larger number of training examples used to build the Twitter doc2vec model. While word vectors are shared between documents, document vectors are always unique in each new document (Le and Mikolov, 2014). Compared to our Reddit corpus, this Twitter corpus includes a higher number of training documents but each document is shorter in length. Thus using the Twitter corpus in training vector representations may increase the complexity of the doc2vec model more than the word2vec model. To achieve higher accuracy while maintaining consistency, we hereafter use training examples to build models for feature vector generation.

| | SVM | | NN | |
|----------|--------|---------|--------|---------|
| | Reddit | Twitter | Reddit | Twitter |
| word2vec | 0.906 | 0.813 | 0.900 | 0.786 |
| doc2vec | 0.772 | 0.803 | 0.797 | 0.823 |
| LDA | 0.868 | 0.748 | 0.846 | 0.721 |

Table 4: Accuracies from feature vectors generated by models trained with Reddit data vs Twitter data.

5 Results

Several quantitative results are discussed, below.

5.1 Frequency

To compare differences in lexicon, we use the entire labelled data set of 22,808 Reddit posts. We compute the frequencies of all unigrams over both the Anxiety and Control sets. The top 200 unigrams for each category are sorted, and the unigrams which appear in both lists are removed, in order to find differentiating subsets. We use the same process for finding the most frequent bigrams. Table 5 summarizes the top 15 most frequent unigrams and top 10 most frequent bigrams, for each group.

| Category | most frequent n-grams |
|-------------------------|--|
| Anxiety unigrams | <i>anxiety, myself, anyone, social, panic, friends, feeling, having, anxious, else, talk, bad, thought, better, felt</i> |
| Control unigrams | <i>our, call, us, edit, old, tell, phone, use, give, same, customer, post, money, let, reddit</i> |
| Anxiety bigrams | <i>(my anxiety), (social anxiety), (my life), (anxiety and), (anxiety i), (anyone else), (talk to), (panic attacks), (panic attack), (where i)</i> |
| Control bigrams | <i>(we are), (from the), (we have), (she was), (he was), (thank you), (that the), (and he), (and she), (what is)</i> |

Table 5: Frequent N -grams ($N=1, 2$) in each class.

The Anxiety unigrams and bigrams explicitly mention anxiety and anxiety-related conditions such as social anxiety and panic attacks. Among the most frequent Anxiety unigrams are words related to feelings (e.g., *feeling, thought, felt, bad*). In contrast, unigrams and bigrams in Control data contain vocabulary general to Reddit (e.g., *edit, post, Reddit*). Control group data from `r/talesfromcallcenters`, `r/talesfromtechsupport` and `r/talesfromretail` contain unique customer- and phone-related words (e.g., *call, phone, customer*) that are not frequently present in the Anxiety group data. The Control set also frequently contains more third-person and first-person plural pronouns compared than the Anxiety set. The most frequent unigrams and bigrams of the Anxiety set include more first-person singular pronouns, however.

5.2 Collocations

Studying collocations captures how groups of words are combined to produce meaning beyond the sum of individual component words. While N -gram frequencies in the previous section reveal how often words appear, identifying collocations can reveal important topics mentioned within a corpus. To find the collocations in both the Control and Anxiety posts, we again analyze the entire data set. Using the NLTK collocation library, we filter collocations by empirically selecting a minimum frequency of 100 for bigrams and 75 for trigrams. We then extract the 30 most collocated N -grams ranked by pointwise mutual information (Manning et al., 1999) from each of the Anxiety and Control sets. We also remove collocations that appear in both the Anxiety and Control collocation lists. Table 6 summarizes the top 10 most collocated bigrams and trigrams for both groups.

| Category | most collocated N -grams |
|-------------------------|--|
| Anxiety bigrams | <i>(self esteem), (side effects), (mental illness), (heart rate), (x post), (mental health), (physical symptoms), (social media), (panic attack), (hey guys)</i> |
| Control bigrams | <i>(blah blah), (front page), (difference between), (tech support), (credit card), (cannot afford), (weeks ago), (customer service), (minutes later), (last night)</i> |
| Anxiety trigrams | <i>(does anyone else), (thanks for reading), (no matter how), (wondering if anyone), (having panic attacks), (stop thinking about), (in high school), (get rid of), (has anyone else), (wanted to share)</i> |
| Control trigrams | <i>(thanks in advance), (the difference between), (the front page), (my best friend), (take care of), (at this point), (a call center), (as far as), (a few days), (a few minutes)</i> |

Table 6: Highly collocated N -grams ($N=2, 3$).

Both bigram and trigram collocations in the Control group show timestamps (e.g., *last night, weeks ago, minutes later, a few days, a few minutes*). Members of the Anxiety community share self-esteem issues, side effects of drugs, how their lives interact with social media, and the physical symptoms of their experiences. Trigram collocations in the Anxiety set are predominantly phrases to ask for advice and find people with the common experiences (e.g., *does anyone else, wonder-*

ing if anyone, has anyone else, wanted to share). There are also collocations that indicate age information (e.g., *in high school*), and users' struggles with anxiety disorders (e.g., *no matter how, stop thinking about, get rid of*).

5.3 Classification

Table 7 summarizes the 10-fold cross-validated accuracy and precision rates of using various types of feature, across logistic regression (LR), a linear kernel support vector machine (SVM), and a neural network (NN) for binary classification. The LR and SVM classifiers were implemented with SciKit-Learn (Pedregosa et al., 2011). We built a custom 2-layer neural network with 256 hidden units per layer and sigmoid activations. During optimization, we empirically used a batch size of 500 and a learning rate of 0.01 for 200 iterations.

Overall, all features are useful in classifying anxiety-related posts on Reddit. For single-source features, we achieve the best results, of 91% accuracy, through word-vector embeddings (word2vec), and through N -gram probabilities. The performance of word2vec is slightly better than the word-vector techniques used by Preotiuc-Pietro et al. (2015) on the Coppersmith Twitter corpus (Coppersmith et al., 2015). By contrast, using N -gram probabilities achieve an overall slightly better precision (92% with NN) than word2vec (91% with SVM). The LDA topic features also perform better than previous results using LDA to detect depression on Twitter (Resnik et al., 2015). Whether topic modelling is more appropriate for long-form posts, as in our data, is the subject of future work.

Since our data did not include meta-data, we implemented content-based features from De Choudhury et al. (2013) including emotion, linguistic style (from LIWC 2007), and an anxiety lexicon. In addition, we combined LIWC and LDA features from Nguyen et al. (2014). The accuracies and precisions of these implementations, as well as the aggregate features, are summarized in Table 8.

For combined methods, our neural network classifier consistently produces the best results. We achieve the highest of accuracy of 98% by combining LIWC with N -gram probabilities and by combining word-vector embeddings (word2vec) with LIWC using this classifier. We improve classification accuracy by 7% over only

using word2vec and by 13% over the LIWC-only baseline. Also, N -grams+LIWC (99%) achieves slightly higher precision than word2vec+LIWC (98%), which is consistent with the difference in N -grams-only word2vec-only results. Combined models, specifically word2vec+ N -gram probabilities, word2vec+LDA, and LIWC+LDA (Nguyen et al., 2014), achieve comparable results with 95%, 94%, and 95% precision, respectively.

For all accuracy and precision values in Table 7 and Table 8, the associated recall was high; between 79% and 99% depending on the classifier. The neural network classifier consistently produced recall values above 90% with variances in the order of 10^{-4} . The SVM classifier produced the lowest recall (79%-90%) with larger variances in the order of 10^{-2} . This fluctuation may be due to using a linear kernel which has a lower representational power than a non-linear kernel

6 Discussion

The LIWC 2015 dictionary provides sufficient coverage of anxiety-related word usage to successfully classify Anxiety and Control Reddit posts. However, by combining LIWC features with N -gram probabilities or unsupervised feature-generation techniques (i.e., vector space embeddings and LDA Topic modeling), we can elevate the classification accuracy to 98%. Moreover, we find correlations between anxiety and specific LDA topics such as school and alcohol (and drug) consumption (see Table 2). This could be an effective method of identifying topics that people with anxiety or other mental illnesses discuss online. By counting unigram and bigram frequency, we also find lexicons relating to feelings and first-person, singular pronouns predominantly represented in the Anxiety group. Furthermore, studying frequent collocations suggests that authors of anxiety-related posts are looking to find other people sharing similar experiences with anxiety.

Due to the relatively recent popularity in the platform, little work has involved the linguistic aspects of Reddit, compared to Twitter. The lengths of posts and community organization of the website suggests considerable potential for sophisticated methods of feature extraction as well as qualitative analysis.

Despite the wide prevalence of anxiety disorders, few attempts have been made to create models capable of automatically detecting the disorder.

| Feature | LR | | SVM | | NN | |
|-----------------|---------------|---------------|----------------------|----------------------|----------------------|----------------------|
| | Accuracy | Precision | Accuracy | Precision | Accuracy | Precision |
| word2vec | 0.87 (9.3E-5) | 0.89 (3.4E-4) | 0.91 (1.0E-5) | 0.91 (7.8E-5) | 0.90 (9.1E-5) | 0.91 (2.2E-3) |
| doc2vec | 0.78 (9.4E-5) | 0.75 (3.7E-4) | 0.77 (8.0E-6) | 0.76 (3.0E-5) | 0.80 (9.5E-5) | 0.78 (2.3E-4) |
| LDA | 0.80 (9.8E-5) | 0.80 (1.6E-3) | 0.87 (1.4E-4) | 0.87 (2.8E-4) | 0.85 (1.6E-4) | 0.85 (7.5E-4) |
| LIWC | 0.85 (4.2E-4) | 0.85 (2.1E-3) | 0.71 (1.1E-2) | 0.81 (1.5E-2) | 0.82 (1.0E-2) | 0.85 (4.9E-3) |
| <i>N</i> -grams | 0.90 (8.8E-5) | 0.89 (4.1E-4) | 0.85 (6.9E-3) | 0.86 (9.4E-3) | 0.91 (2.2E-4) | 0.92 (1.4E-3) |

Table 7: Accuracies, precisions (and variances) for single-source features.

| Feature | LR | | SVM | | NN | |
|----------------------------|---------------|---------------|---------------|---------------|----------------------|----------------------|
| | Accuracy | Precision | Accuracy | Precision | Accuracy | Precision |
| word2vec + LIWC | 0.89 (7.9E-4) | 0.90 (7.3E-4) | 0.84 (2.8E-3) | 0.88 (1.2E-2) | 0.98 (2.3E-6) | 0.98 (1.7E-5) |
| doc2vec + LIWC | 0.84 (9.3E-4) | 0.88 (1.4E-3) | 0.80 (3.7E-3) | 0.87 (1.1E-2) | 0.92 (2.4E-5) | 0.93 (1.3E-4) |
| word2vec + LDA | 0.87 (3.0E-4) | 0.89 (5.5E-4) | 0.91 (2.1E-5) | 0.91 (1.6E-4) | 0.92 (1.8E-5) | 0.94 (8.8E-5) |
| doc2vec + LDA | 0.82 (5.4E-4) | 0.85 (6.9E-4) | 0.82 (2.8E-3) | 0.83 (2.7E-4) | 0.84 (8.2E-5) | 0.86 (8.7E-5) |
| word2vec + <i>N</i> -grams | 0.90 (2.5E-5) | 0.88 (2.1E-4) | 0.89 (1.5E-3) | 0.91 (2.2E-3) | 0.91 (2.4E-4) | 0.94 (7.2E-4) |
| doc2vec + <i>N</i> -grams | 0.90 (3.5E-4) | 0.90 (8.0E-4) | 0.74 (3.4E-3) | 0.90 (1.7E-2) | 0.92 (7.5E-5) | 0.93 (8.6E-4) |
| LDA + <i>N</i> -grams | 0.90 (1.3E-4) | 0.88 (9.6E-4) | 0.88 (2.9E-3) | 0.89 (8.3E-3) | 0.90 (1.3E-3) | 0.93 (1.1E-3) |
| LIWC + <i>N</i> -gram | 0.92 (6.2E-5) | 0.91 (2.6E-4) | 0.98 (2.4E-4) | 0.91 (2.9E-4) | 0.98 (2.0E-5) | 0.99 (1.1E-5) |
| Nguyen et al. | 0.81 (6.0E-3) | 0.86 (4.2E-3) | 0.81 (3.4E-3) | 0.79 (7.0E-3) | 0.93 (1.2E-4) | 0.94 (7.4E-4) |
| De Choudhury et al. | 0.64 (8.7E-4) | 0.67 (4.8E-3) | 0.62 (4.7E-3) | 0.60 (5.1E-3) | 0.85 (1.4E-3) | 0.88 (6.3E-3) |

Table 8: Accuracies, precisions (and variances) for aggregate features.

Further work should also include larger data sets in combination with explicitly associated diagnostic criteria, assessments, or health records, to emphasize validity.

References

- American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition: DSM-5®*. American Psychiatric Association.
- Steven Bird. 2006. Nltk: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*. Association for Computational Linguistics, pages 69–72.
- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3(Jan):993–1022.
- Glen Coppersmith, Mark Dredze, Craig Harman, Kristy Hollingshead, and Margaret Mitchell. 2015. CLPsych 2015 shared task: Depression and PTSD on Twitter. In *Proceedings of ACL*. pages 31–39.
- Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on Reddit: Self-disclosure, social support, and anonymity. In *Proceedings of ICWSM*. pages 71–80.
- Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Proceedings of ICWSM*. page 2.
- Koen Demyttenaere, Ronny Bruffaerts, Jose Posada-Villa, Isabelle Gasquet, Viviane Kovess, Jean Pierre Lepine, et al. 2004. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health surveys. *Jama* 291(21):2581–2590.
- Katya C Fernandez, Cheri A Levinson, and Thomas L Rodebaugh. 2012. Profiling: Predicting social anxiety from Facebook profiles. *Social Psychological and Personality Science* 3(6):706–713.
- Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford* 1(12).
- Mrinal Kumar, Mark Dredze, Glen Coppersmith, and Munmun De Choudhury. 2015. Detecting changes in suicide content manifested in social media following celebrity suicides. In *Proceedings of ACM*. ACM, pages 85–94.
- Quoc V Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *Proceedings of ICML*. volume 14, pages 1188–1196.
- Christopher D Manning, Hinrich Schütze, et al. 1999. *Foundations of statistical natural language processing*, volume 999. MIT Press.
- Kathleen Ries Merikangas, Jian-ping He, Marcy Burstein, Sonja A Swanson, Shelli Avenevoli, Lihong Cui, et al. 2010. Lifetime prevalence of mental disorders in us adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry* 49(10):980–989.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word

representations in vector space. *arXiv preprint arXiv:1301.3781*.

Moin Nadeem. 2016. [Identifying depression on Twitter](#). *CoRR* abs/1607.07384. <http://arxiv.org/abs/1607.07384>.

Thin Nguyen, Dinh Phung, Bo Dao, Svetha Venkatesh, and Michael Berk. 2014. Affective and content analysis of online depression communities. *IEEE Transactions on Affective Computing* 5(3):217–226.

Ted Pedersen. 2015. Screening Twitter users for depression and PTSD with lexical decision lists. In *Proceedings of ACL*. pages 46–53.

Fabian Pedregosa, Gael Varoquaux, Alexander Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015. Technical report, University of Texas at Austin, Austin, Texas.

Daniel Preotiuc-Pietro, Maarten Sap, H Andrew Schwartz, and LH Ungar. 2015. Mental illness detection at the world well-being project for the clpsych 2015 shared task. *Proceedings of NAACL HLT 2015* page 40.

Philip Resnik, William Armstrong, Leonardo Claudino, Thang Nguyen, Viet-An Nguyen, and Jordan Boyd-Graber. 2015. Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. In *Proceedings of ACL*. pages 99–107.

Kathleen C Thomas, Alan R Ellis, Thomas R Konrad, Charles E Holzer, and Joseph P Morrissey. 2009. County-level estimates of mental health professional shortage in the United States. *Psychiatric Services* 60(10):1323–1328.