Sentiment Analysis with Twitter Data:

Identifying Depression Using NLP

Malik Hasan and Diego Gomez

San Jose State University

ENGR 195 E

Abstract

Natural language processing and sentiment analysis are up-and-coming in the tech world, and depression is also a growing problem in the modern world. This study aims to use technological methods in order to address the depression epidemic using the vast amount of personal data available in social media posts, specifically, Twitter. Using sentiment analysis and analyzing other factors such as posting times, this study uses machine learning algorithms like Naive Bayes, Logistic Regression and Neural Networks, to see if we can produce an effective model to detect depressive symptoms and evaluate which model is most effective and which factors should be weighted more. We found logistic regression models worked best for identifying tweet sentiment and evaluating users for low average tweet sentiment and tweet frequency identifies depression well. The findings from this study can be used for further psychological research into depression as well as early intervention and industrial/clinical uses.

**Introduction**

Depression rates have grown exponentially in our age, and relatively little funding or research is diverted towards investigating and treating it (Hidaka, 2012). For that reason, our goal was to research a simple cost-effective way to approach early intervention and suicide prevention. In this study, we try to tackle the growth of depression by using machine learning and natural language processing methods to identify at-risk individuals online.

**Sentiment Analysis**

Sentiment analysis is defined as the process of mining subjective text and extracting the social sentiment, whether it be positive, neutral, or negative. Sentiment analysis is also a subfield of Natural Language Processing, which involves computer understanding of human language. In recent years, sentiment analysis has improved from accurately identifying the sentiment at the sentence level to identifying the sentiment at the individual phrase level (Agarwal, Vovsha, Rambow, & Passonneau, 2011).

Sentiment analysis has gained popularity over time. One of the main reasons for this increased interest is the increased popularity of social media applications, which provide a rich resource for text mining large data sets. Of the various applications used for sentiment analysis, Twitter is the most common. The main reason is that Twitter only allows a limited number of characters per tweet. This forces users to get their point across with brevity unlike other platforms, such as Facebook, where users can be more verbose if they choose to.

Research on sentiment analysis using Twitter has utilized interesting methods to analyze the sentiment of the posts. For example, Go, Bhayani, and Huang (2009) and Bermingham and Smeaton (2010) used sentiment analysis on Twitter by using tweets ending in negative emoticons as negative and tweets ending in positive emotions as positive. In these studies, the authors built models using Naive Bayes and Support Vector Machines, which they showed outperform other classifiers. The results from these studies were promising and provide a strong point of origin for continued research in this domain.

**Depression**

Depression is a form of mental illness and is defined as the onset of low self-esteem, negative mood, decrease in interest in pursuing activities that elicit pleasure, and lowered overall experience of pleasure (De Choudhury, Gamon, Counts, & Horvitz, 2013). Furthermore, individuals with depression focus on negative and unhappy information and have an overall negative perspective on life (Kessler, Berglund, Demler 2003; Rude, Gortner, & Pennebaker 2004).

It is difficult to detect the precursor signs of depression (Paul & Dredze, 2011), but recent research suggests that social media postings may be an effective method of identifying individuals at risk (De Choudhury, et al., 2013). In this study, researchers build a statistical classifier to identify those at risk of depression, before onset. Results indicated that a decreased use of social media and an increased spread of content with negative affect (among other signals) were primary characteristics of identifying individuals at risk of depression (Park et al., 2012). This study suggests that social media can be leveraged to provide early intervention to individuals at risk of depression.

#### **Recent Research**

Wang, Zhang, and Sun (2013) conducted a study on depression in Chinese micro-blogs. Although the study was conducted on Chinese text, a lot of the process is applicable to English analysis. Some insights can be very useful for our project such as the idea that “If a micro-blog content is long, the beginning and ending sub-sentences often reflect the writer’s feelings more directly, thus more important than those in the middle” (Wang et al., 2013).

Furthermore, they make it clear that determining that the sentimental value of an individual’s micro-blogs or Tweets is very negative is not enough to determine depression. There are a number of other factors at play that can help identify depression rather than simple negative messaging. Other factors that they took into account based on previous psychological literature include: 1. Depressed individuals use the first person singular pronouns (I, me, myself) more often, and use the first person plural pronoun (we, us, ourselves) less often than usual; 2. They use far less emojis; 3. They are much more likely to suffer from chronic insomnia and post on social media more often than usual between midnight and 6 am; 4. They post more original posts than usual rather than reTweets (Wang et al., 2013).

**Study Aim**

As machine learning becomes more advanced and more relevant in the current world, we find new questions for machine learning to solve. Today we ask is it possible to identify Tweets with a negative or depressive connotation using Machine Learning? Could we identify people at risk of depression using this technology? If so, the applications of such algorithms are many and can be used as a method to identify individuals with mental health disorders, apply early interventions, or gain interest for further research.

**Methods**

The study was divided into two stages. First, a supervised learning stage was necessary to train models to recognize positive and negative sentiment in tweets. Then once the tweets had been properly labeled by the best model, the next step was to use statistical methods to identify users at risk of depression based on the sentiment of that user’s tweets.

**Data**

To this end, we used a dataset containing 1.6 million tweets labeled with their sentiment value from Stanford. These tweets were labeled 0 (negative), 2 (neutral), or 4 (positive), however very few of the tweets were labeled neutral so we omitted them from many of the analyses. These tweets averaged 14-15 words in length. There were 2-3 tweets per user in the sample on average, but this distribution was skewed and ranged up to 550 tweets from a single user.

In order to process the text of the tweets, we used the NLTK casual tokenizer to parse the words. This tokenizer was designed by Christopher Potts at Stanford to be a “Twitter-aware tokenizer, designed to be flexible and easy to adapt to new domains and tasks,” and for our purposes seemed to work very well. Furthermore, we noticed many of the tweets contained references to other users which were not relevant to our purposes so we removed any word where the first character was ‘@’ in order to avoid swamping the models with user tags.

Another issue that we dealt with was the presence of some users who clearly resembled bots and were not relevant for our purposes. These were identified easily because they tweeted the same text repeatedly. We cleaned this spam when necessary by ignoring users with repetitive tweets and few unique tweets.

**Results**

**Supervised Stage**

In the first stage, we used the labeled data to train various models to identify the sentiment of a tweet and evaluated which worked the best for our problem.

**Naive Bayes.**

We first constructed from scratch a Naive Bayes model to approach this problem. Using a bag of words method and add-1 smoothing to account for new words, this simple model yielded a 78% accuracy score. Naive Bayes is a decent and simple way to evaluate tweet sentiment.

**Logistic Regression.**

In the Logistic Regression model, we used Cross-validation to allocate the best parameter estimates that minimized error. With the default number iterations (100), we were able to achieve 78% accuracy. However, Python recommended increasing the number of iterations to reach the global minimum. We changed the number of iterations to 500. In this model, the algorithm achieved a 79% accuracy rate.

**SVM.**

The Support Vector Machine (SVM) did not converge to a global minimum. As such, the algorithm ran indefinitely. Although the parameters were changed to modify the algorithm using iteration, the algorithm was not functional with the data set.

**KNN.**

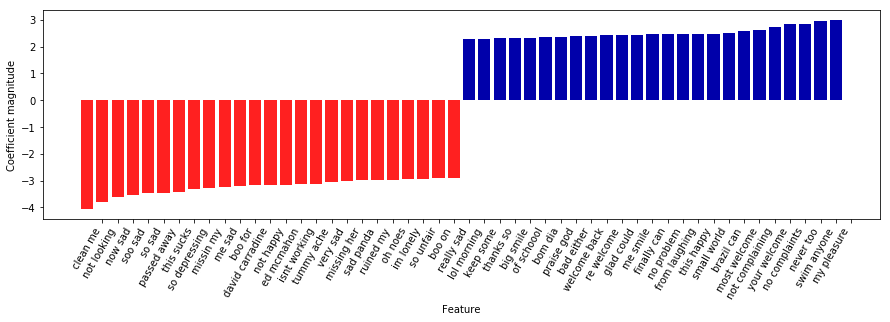
Similar to SVM. the K-nearest neighbors algorithm did not converge to a global minimum and ran indefinitely. We used iteration to modify the *k* parameter, but ultimately it did not produce any results.

**RNN.**

Using the Pytorch library, we implemented a recurrent neural network by hand. We thought that a Recurrent Neural Network (RNN) would be suited for a classification problem like the one at hand, however, it proved not to be the case. After multiple attempts at training for multiple days with slightly different hyperparameters, the neural network still failed to discern positive and negative tweets; we were forced to move on from this model. We also attempted to use a neural network from different libraries but met with similarly disappointing results. It seems that the model is not suitable for the data or that our computers lacked the resources to train it properly.

**Top Negative/Positive Features**

Using the “mglearn” library in Python, it was possible to plot the top 25 negative and top 25 positive two word phrases in the data set, based on coefficient magnitudes.

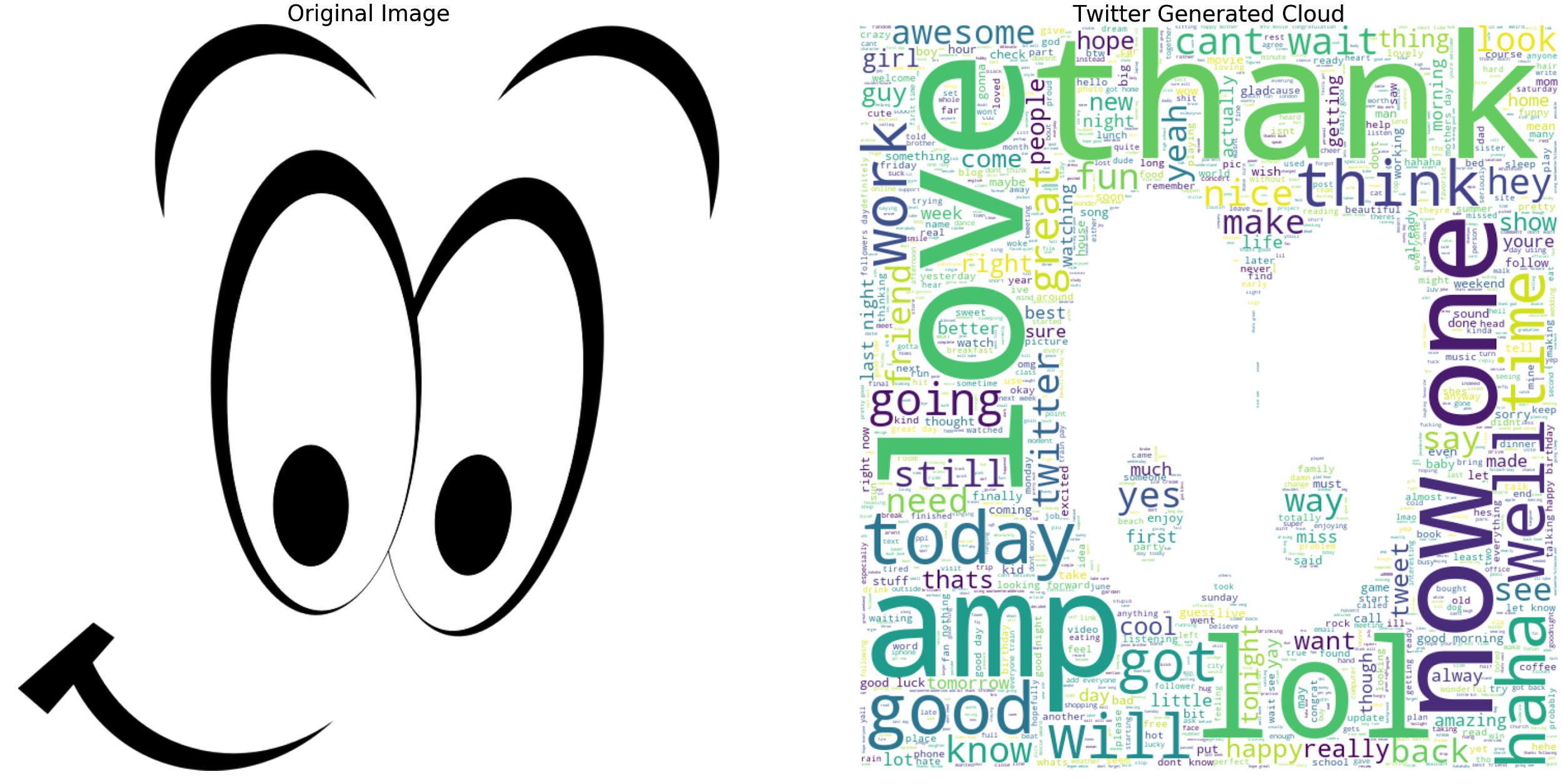


It is evident that phrases such as “so sad”, “me sad”, and “im lonely” convey an extremely negative sentiment. It is also evident that phrases such as “big smile” and “me smile” convey positive sentiment. It is noteworthy to mention that the context is missing. Specifically, it is not clear if these phrases were used in a sarcastic tone, in which case, the sentiment may have been erroneously labeled. This presents one of the most prominent issues with sentiment analysis--the fact that sarcasm and double negatives are difficult to identify. This issue is magnified in a social media setting such as Twitter, where sarcasm is more likely than other social media outlets, such as LinkedIn, for example.

**Wordcloud**

To have visuals aid our study, we created a word cloud for both negative and positive tweets.

**Positive Tweets Wordcloud.**



As visible from the image above, we used a mask for the positive tweets word cloud. It was suggested in the documentation to have an image with a white background. Unfortunately, the complete image was not recognized. Nonetheless, we can see that words like “love”, “good”, “well” and “haha” were among the most common words used that were labeled with positive sentiment.

**Negative Tweets Wordcloud.**



In the negative tweets word cloud, the mask was not picked up at all. Regardless, we can see that words such as “sad”, “hate”, and “miss” were among the most common words labeled with negative sentiment. Interestingly, words such as “love” are present, perhaps they were used in juxtaposition to other negative words to show the contrasting effect from users.

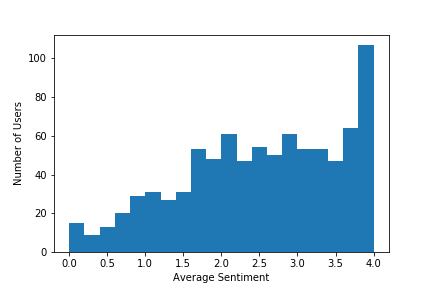
**SQL**

We had an account of the DB2 database server from IBM, where we can use SQL. We were able to use the credentials and establish a connection to the server. We were able to use the ibm\_db\_dbi library to run SQL in Python. The following line of code (*selectQuery = "select \* from TWEETS"*) allowed us to extract the data set. From this query, we were able to use pandas to convert the query into a data frame. Unfortunately, the database server was not successful at detecting specific columns from the data set. This prevented us from running SQL queries to extract meaningful information using aggregate functions. After hours spent troubleshooting, we confirmed that the issue was due to not having a paid account, which prevents users of the free version of the service from having multiple schemas.

**Testing for Depression**

**Statistical Tests.**

In order to test whether there was any discernible difference in individuals that could be identified from the sentiment values, we ran some statistical tests. The following histogram plots the average sentiment of the set of tweets for each of the 873 users with at least 50 tweets in the dataset.



It is clear that the average sentiment of the users is highly skewed left, and the majority of users tend toward positive tweets. This is further evidenced by running a one-sample z-test comparing the given data to the hypothetical normal distribution which would be expected if the users tended to tweet positively and negatively at similar rates. The p-value is 4.43 x 10-48 which is more than enough evidence that users favor positive tweets overall.

Furthermore, less than 10% of the users had average sentiments lower than 1. Therefore the users who had overwhelmingly negative tweets deviated significantly from the norm and provide a reasonable basis from which we can look for traits and traces of depression and mental illness. This is also based on the findings of Park et al. (2012) who found that depressed users of social media post a higher proportion of negative content.

It is furthermore evidenced by case studies of users with very low average sentiment. Here is a sample of tweets from a single user who has overwhelmingly negative tweets overall.

@travisking i'm such a fattie! xD i want food in general..well...mostly pancakes.

@NoRaptors What? No...it's not for fun anymore. I don't cut for fun. It was a joke... Agh. Nvm. -holds self-

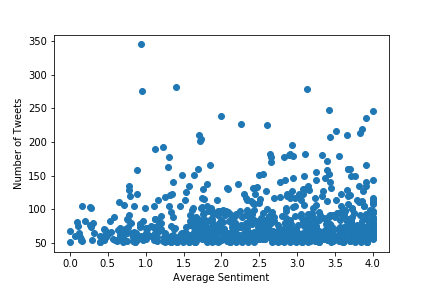
Is anyone here?? Lori? Paige?

I wish Jess was awake. Argh... I miss you Jess. -bangs head against wall-

Jess please wake up. Please. I need you. #thisismewhining #imissjess #sorryyouhaftalisten #turningintotravis

Even from this small sample of the user’s tweets, signs of depression are evident. This suggests that analyzing the average sentiment of a user’s tweets is adequate at identifying individuals that may be at risk of depression.

Based on these calculations, we hypothesized that users that had an average sentiment below 0.5 which is two standard deviations below the mean at 2.5 were showing signs of depression. This hypothesis was further corroborated and narrowed down when we compared the average number of tweets for users below and above the threshold. According to Wang et al. (2013) and Park et al. (2012) depressed users tweet less often. Our data showed the most significant drop in tweet frequency below a threshold of 0.7 average sentiment. Using an ANOVA test comparing tweet frequency above and below that threshold, we obtained a p-value less than 0.01indicating that there is indeed a statistically significant difference in tweet frequency.



Based on these findings we adjusted our threshold between normal and potentially depressed users at 0.7 average sentiments. Wang et al. also describe a difference in the times that depressed users tweet, tweeting more often between midnight and 6 a.m. (2013). We analyzed the data using an ANOVA test comparing the tweeting times above and below our threshold but found no significant difference. Future studies should look into this aspect deeper to determine whether it is indeed a factor.

**Conclusions**

In conclusion, we found that the best model for evaluating tweet sentiment was a logistic regression model. Once tweets have been labeled by the model, depression can be reasonably identified by analyzing the average sentiment and frequency of a user’s tweets. Users with an average sentiment below 0.7 on our scale of 0-4 and a low tweet frequency are the most likely to show signs of depression and should be the focus of our medical and psychological resources. This is consistent with most of the findings of Park et al. (2012) and Wang et al. (2013).

**Limitations**

Working with a data set of this magnitude presented a wide array of limitations. For instance, model training was a process that would take several hours, and in the case of neural networks, the process took multiple days. Another limitation was that we could not run SQL through the DB2 database from IBM. Although we did get a connection to the database, it did not detect the columns from our dataset. After time spent troubleshooting, it was identified that DB2 did not detect columns because an upgrade was necessary to provide multiple schemas. However, financial resources were necessary to upgrade.

References

Adam Bermingham and Alan Smeaton. 2010. Classifying sentiment in microblogs: is brevity an advantage is brevity an advantage? ACM, pages 1833–1836.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. J. (2011). Sentiment analysis of twitter data. In Proceedings of the Workshop on Language in Social Media (LSM 2011) (pp. 30-38).

Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. Technical report, Stanford.

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013, June). Predicting depression via social media. In Seventh International AAAI conference on weblogs and social media.

Detels, R. 2009.The scope and concerns of public health. Oxford University Press.

Hidaka, B. (2012). Depression as a disease of modernity: Explanations for increasing prevalence. Journal of Affective Disorders, 140(3), 205-214.

Kessler, R.C., Berglund, P., Demler, O. et al. 2003. The Epidemiology of Major Depressive Disorder: Results from the National Comorbidity Survey Replication. Journal of the American Medical Association 289 (23): 3095-3105.

Park, M.; Cha, C.; & Cha, M. 2012. Depressive Moods of Users Captured on Twitter. In Proc. ACM SIGKDD Workshop on Healthcare Informatics (HI-KDD).

Paul, M., J.; & Dredze, M. 2011. You are What You Tweet: Analyzing Twitter for Public Health. In Proc. ICWSM ’11.

Rude, S. Gortner, E. & Pennebaker, J. 2004. Language use of depressed and depression-vulnerable college students. Cognition and Emotion, 1121-1133.

Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. 2015. Recognizing Depression from Twitter Activity. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI ’15). Association for Computing Machinery, New York, NY, USA, 3187–3196. DOI:https://doi.org/10.1145/2702123.2702280

Wang, X., Zhang, C., & Sun, L. (2013). An improved model for depression detection in micro-blog social network doi:10.1109/ICDMW.2013.132