Predicting Depression in Social Media Posts Using Natural Language Processing

Corpus Analysis Spring 2021
Andrew Kirby

Abstract

Since the onset of COVID-19, people have turned to social networks as a replacement for interaction and interpersonal self-expression that had once taken place offline. While it is possible for a human to easily detect the mood of a speaker or writer, doing so can pose a challenge for a computer. It is possible for a computer program to distinguish between a text written by a depressed person and a text written by a non-depressed person. Developing tools which can successfully discern a writer's mood can be a useful tool for assisting mental health professionals in their work, be it in a corporate environment or beyond.

1 Introduction

In 2012, the average person spent approximately 90 minutes per day on social media; as of 2020, that number has increased to 145 (Tankovska 2021). The increasing time investment in social media activity is an illustration of its growing relevance in the average person's life. Consequently, its reflection is becoming ever more meaningful, with posts and tweets being recorded for better or for worse. While psychology in the past was tasked with examining a person's behavior in real-time, more of this behavior is now taking place on the internet. Currently, psychologists favor an approach called Cognitive Behavioral Therapy (CBT), which favors targeting unhelpful thoughts or learned patterns of behavior, and assisting the person with learning new, helpful patterns (APA 2017). One major application that CBT has involves treating depression. Signs of depression can include persistent sadness, feelings of guilt or loss, and problems with sleeping, concentration, or appetite (NIMH 2015).

There are companies which intend to supplement (or possibly replace) the psychologist. One such example is WoeBot, which claims in a study that users can form a 'therapeutic bond' with its AI in a way that can help persons dealing with a trauma. WoeBot is a conversational agent that engages the user with exercises and gives 'empathic statements', along with other protocols, that help the user (Darcy et al., 2021). In order to respond to a person in an empathetic way, the AI will need to be able to discern the mood of the person chatting with it. Natural Language Processing (NLP) techniques are required to accomplish this task.

2 Text Processing Techniques

Common approaches of processing text, changing it into numbers that a program can understand, consist of using Bag of Words (BoW) and TF-IDF (Term Frequency-Inverse Document Frequency). The Bag of Words approach consists of counting up all the words and putting them into vectors so that they can be computed into other forms of information. TF-IDF takes this a step further, by giving less importance to words that occur very frequently while not having a lot of useful content, like 'the' and 'a' or 'it'. It also takes into account the quantity of documents in a document set in order to determine which words get more value (Eisenstein 2019, Ch. 2). The TF-IDF approach was introduced by Spärck Jones (1972).

Subjectivity measurements involve determining the degree to which a parcel of text is subjective or objective. This area of study is also commonly referred to as opinion mining, and tools available often integrate this feature with sentiment analysis. This is accomplished by using a model to make a prediction on another piece of text (Murray & Carenini 2009).

Sentiment Analysis is a measure of how positive or negative a parcel of text is. This can be accomplished in a variety of ways. Some examples include using data like tweets with happy or sad emoticons, product reviews with ratings, or a lexicon with positive and negative words to generate a quantity that corresponds with how positive or negative the item of text is (Eisenstein 2019, Ch. 4). Sentiment analysis can also be approached as a text classification problem, where machine algorithms are put to work to detect positive or negative sentiment. For example, film review data can be used to train a model to predict sentiment of other pieces of text (Pang & Lee 2004).

For classification, the Support Vector Machine (SVM) and Logistic Regression (LR) are popular, especially for binary classification. The SVM works by increasing the dimensions of a training vector and then classifying using a line (Henrique et al., 2019). LR is used with binary dependent variables, which have two values as in 0/1 or True/False. As a side effect of classification, it estimates the posterior probability of true and false labels. using a scoring function for base features (like word vectors) and a binary label (Eisenstein 2019, Ch. 2).

3 Past Work

Many systems have been created recently which use social media data to predict the mood of the person posting text online. These systems rely, for the most part, on the basic techniques described above, with more sophisticated ones relying on neural networks or more computationally demanding methods.

One system took approximately 12,000 posts from Reddit, about ⅓ of which were classed as depressed. Two other sets were also included: one as a control, and another consisting of not-depressed posts. This system used the LIWC (Linguistic Inquiry and Word Count) tool, which is based on a notion that the words a person uses reflects their psychological state. For text processing, the authors used a smoothed version of TF-IDF, and SVM for classification (Wolohan et al., 2018).

Another system worked with Arabic language text, collecting its data from an online forum called Nafsany.[1] In data processing, the authors used a tool to remove non-contentful words, normalized morphological characteristics (necessary for the Arabic language), and performed stemming. Sets were split into depressed and not depressed. It then used Arabic lexicons along with n-gram extraction, TF-IDF, and classified with Random Forest (RF), SVM, K-Nearest Neighbors, and Stochastic Gradient Descent (Alghamdi 2020).

A different system used 'depression' and 'neurotic' stream-of-consciousness essays from college students, alongside LIWC. These essays essentially required students to write roughly 750 words about their thoughts of themselves and their positions in existence. This system changed the text into vectors using Latent Dirichlet Allocation (LDA), and classified it with a linear regression (Resnik et al., 2013).

Finally, one system also relied on Reddit posts for its data, classifying based on depressed posts from the depression subforum, and 'standard' posts from other parts of the site. This system also used the LIWC, along with LDA and TF-IDF for text vectors. For classification, the authors used and compared LR, SVM, RF in their work (Tadesse et al., 2019).

4 Methodology

The dataset contains 20000 data points, with 10000 each coming from two different sources. Each dataset was shuffled before partitioning, and shuffled again when forming training and testing sets. This particular number of points was chosen because that was what this computer could tolerate. The first dataset contained posts on the depression subforum at Reddit. This dataset was obtained from Kaggle.[2] The other dataset, with happy posts, was also obtained from Kaggle.[3] In the description, the source states that people somewhere were "asked to write the reason why they were happy." Examples from both datasets are included in the Appendix.

---

[1] Nafsany forum(?) https://www.nafsany.cc/
[2] Kaggle depressed post dataset: Kaggle.com
[3] Kaggle happy post dataset: Kaggle.com

I created text vectors from the sentences surrounding the named entities using the `TfidfTransformer` from scikit-learn.[4] I then concatenated these vectors with the sentiment and subjectivity values acquired above. I used unigram features here, as a default.

For sentiment analysis, I used NLTK's VADER[5] (Valence Aware Dictionary for Sentiment Reasoning) as a model for evaluating sentiment polarity. VADER uses a weighted sentiment lexicon and is tuned for 'microblog-like contexts'. This model is then tested on human-annotated gold standard data to measure effectiveness (Hutto & Gilbert, 2014). For evaluating subjectivity, I used TextBlob.[6] TextBlob evaluates subjectivity using a model trained on annotated movie review data from IMDB.

After these datasets were obtained, some light processing was done to remove newlines (\n) and other unwanted characters. Sentiment and subjectivity values were calculated and added to the conjoined dataframe that contained the 20000 data points. A ratio of 80:20 was used for training and testing, after shuffling the data.

For classification, I used `LogisticRegression` from scikit-learn with the following parameters: L1 penalty, a $C = 1.0$, and the `liblinear` solver, which gave the best results on the data.[7] Posts from the depressed dataset were marked with a 0, because depression tends to entail nothingness. Posts from the happy dataset were marked with a 1, because happiness means being one with the world.

---

[4] TF-IDF: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html
[5] NLTK VADER: https://www.nltk.org/howto/sentiment.html
[6] TextBlob: https://textblob.readthedocs.io/en/dev/
[7] LR @ sklearn: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

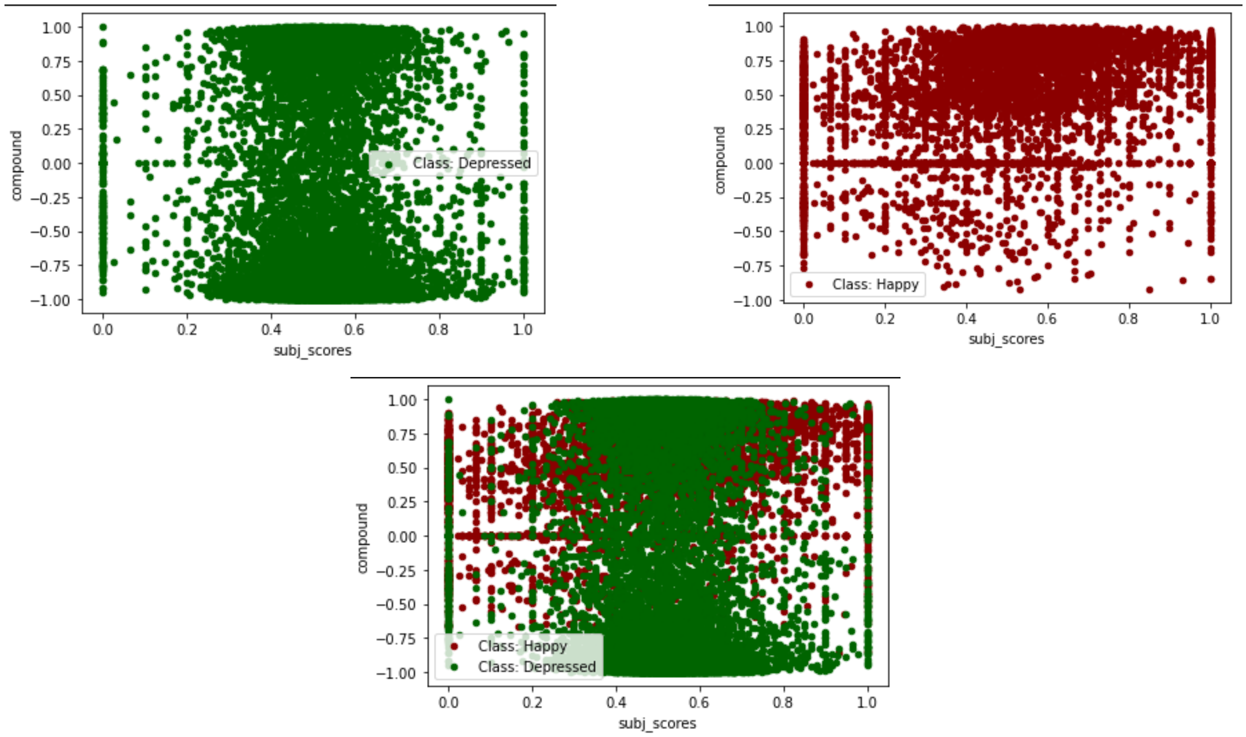

Figure 1: Plots of sentiment and subjectivity analyses on the depressed data (green, top left), and the happy data (red, top right), along with an overlay of the two (bottom, middle).

In Figure 1, there are clear patterns in the data from the sentiment and subjectivity analyses. Depressive posts are moderately subjective, and contain sentiments that skew highly negative or highly positive. Happy posts, meanwhile, skew generally positive, with more of a wider distribution of subjectivity versus objectivity.

5 Results

seed	**Sentiment + Subjectivity: LR**	**TF-IDF: LR**
3	0.72	0.978
4	0.718	0.98
13	0.724	0.979
81	0.707	0.98
333	0.724	0.974
Median:	0.72	0.979

Figure 2: Accuracy scores on the dataset described in the above section. System was run on five different seeds, with the median accuracy at the bottom.

The results show that, while the combination of sentiment and subjectivity scores returned a median accuracy of 0.720 with the LR, the TF-IDF alone with the LR proved to be significantly more accurate, with a median accuracy of 0.979.

6 Discussion

While the initial plots in figure 1 suggested that the sentiment and subjectivity scores correlated with some basic intuitions about the data, these features proved less useful when attempting to make predictions on the nature of the data. Because the TF-IDF matrices were populated with so many values, even when converted to a dense form, concatenating these values with the sentiment and subjectivity values had no effect on the accuracy scores. The computer used in this project did not have sufficient memory to process TF-IDF bigrams for this dataset, so only unigrams were used. This computer also didn’t have enough power to run an SVM on this quantity of data.

7 Conclusion

While the accuracy scores for the sentiment and subjectivity analyses are more or less in line with the reviewed work in this article, the TF-IDF results are exceedingly accurate, suggesting an issue with the data. It may be the case that the datasets are simply too distinct from one another, and that the presence of a neutral control group, rather than a contrastive ‘happy’ dataset may yield more conclusive results. It simply does not make sense that throwing the text into vectors, without doing a lot of the more complex work done (such as with LIWC) in other studies, could outperform them.

For future work, it may be helpful to use human-annotated data, as sometimes people make happy posts in the depression forum, and vice versa. If possible, it would be helpful to gather text produced in some form by a large number of persons who can be confirmed to have been clinically diagnosed with a depressive disorder, for purposes of comparison with each person’s respective analyses. In addition, it might be helpful to allocate or combine more text features, perhaps involving semantic similarity or syntactic features that could be distinct to depressive writers.

Depression is a difficult obstacle to surmount, as it often amounts to a solitary struggle. While technology may prove helpful with identifying and diagnosing depression, it is also important to consider ethical implications of these tools. A person’s privacy and dignity must be respected, and reducing their emotions to a series of values to process risks devaluing that person. It is my hope that, as NLP tools advance, humans will remain sacred above all.

Appendix

Example happy text data	Clas s	Compound Sentiment Score	Subjectivity Score
I can remember many happy events of my life and out of those, I would like to talk about the event that I can still remember vividly regarding my success in the board exam. The moment I heard that I had been awarded a scholarship based on my performance in the board exam, I became the happiest man in the world. This was indeed a very happy moment for me as it is something I was looking forward to achieving and the news made my parents quite happy and proud.	1	0.981	0.75
March 30th was my birthday. Got calls calls and wishes from my friends and relatives whom I didn't expect at all. That made me do happy.	1	0.8126	1
One event that made me happy was purchasing a new baseball game for my Playstation that I could play with my friend.	1	0.8519	0.618182

Example depressed text data	Class	Compound Sentiment Score	Subjectivity Score
Just wondering if anyone else around here has gone through it? I'm currently on Nardil and I don't intend on getting off of it. But the few stories I've found on the withdrawal seem horrible. Not as bad as benzos, but worse than SSRI's/SNRI's. Yet many people also seem to have gone off/on it without any mention of withdrawal?	0	-0.7339	0.544444
A friend with incredibly high IQ, knows why and how things work, doesn't find meaning in life. Hasn't felt happy in years. Hasn't felt sad in years. Knows he is in depression. Doesn't like to talk about it. How can I help?	0	-0.091	0.846667
No matter what, even when surrounded with people that call themselves my friends , even when I have people tell me they like being with me, I can't stop feeling alone. How everyone around me is just lying to me out of pity. All I want is someone to lay my head on and cry to, so they can help me carry the weight I feel. And even though I supposedly already have that, I have only felt a real connection with one person, and I doubt they care about me anymore.	0	0.2551	0.5

Sample text data from each dataset: includes sentiment and subjectivity scores with classification. 0 means depressed, 1 means happy.

References

Alghamdi, N. S., Mahmoud, H. A. H., Abraham, A., Alanazi, S. A., & García-Hernández, L. (2020). Predicting depression symptoms in an Arabic psychological forum. *IEEE Access*, *8*, 57317-57334.

APA. (2017) .What is Cognitive Behavioral Therapy? *American Psychological Association.*

Darcy, A., Daniels, J., Salinger, D., Wicks, P., & Robinson, A. (2021). Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study. *JMIR Formative Research*, *5*(5), e27868.

Eisenstein, J. (2019). *Introduction to Natural Language Processing*. MIT Press.

Hutto, C., & Gilbert, E. (2014, May). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 8, No. 1).

Murray, G., & Carenini, G. (2009). Predicting subjectivity in multimodal conversations. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 1348-1357).

U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Mental Health. (2015). *NIMH Strategic Plan for Research* (NIH Publication No. 02-2650).

Resnik, P., Garron, A., & Resnik, R. (2013, October). Using topic modeling to improve prediction of neuroticism and depression in college students. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1348-1353).

Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation, 28,* 11–21.

Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of depression-related posts in reddit social media forum. *IEEE Access*, *7*, 44883-44893.

Tankovska, H. (2021). Daily social media usage worldwide 2012-2020. *Statista*.

Wolohan, J. T., Hiraga, M., Mukherjee, A., Sayyed, Z. A., & Millard, M. (2018, August). Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with NLP. In *Proceedings of the First International Workshop on Language Cognition and Computational Models* (pp. 11-21).