Detecting Risk for Depression Via Tweets

Alex Reneau
alexreneau2021@u.northwestern.edu
Northwestern University

Ava Robinson
avarobinson2021@u.northwestern.edu
Northwestern University

Gabrielle Klein
gabrielleklein2021@u.northwestern.edu
Northwestern University

Laura Wiseman
laurawiseman2021@u.northwestern.edu
Northwestern University

Rachael Tang
rachaeltang2021@u.northwestern.edu
Northwestern University

ACM Reference Format:
Alex Reneau, Ava Robinson, Gabrielle Klein, Laura Wiseman, and Rachael Tang. 2020. Detecting Risk for Depression Via Tweets. In . ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION, BACKGROUND, AND SIGNIFICANCE

Mental health is one of the most important factors in our daily lives; our thoughts, behaviors, and emotions all depend on our emotional well-being. Sadly, mental illness is extremely common- 46.4% of the United States population will have one or more diagnosed mental disorders at some point in their lives [5]. Moreover, the lifetime prevalence of major depressive disorder in the US is 17%, making it the most common individual mental disorder in the country [5].

However, even with the acknowledgement that mental health is a prevalent issue in our society and the decrease of social stigma [4], it is incredibly difficult for individuals to receive help. Approximately "60 percent of youth with major depression did not receive any mental health treatment in 2017-2018" [3].

Given the large prevalence of mental disorders and difficulty of receiving a diagnosis and treatment, we have decided to build an interactive tool which predicts an individual's risk for depression based on their Twitter data. As American society becomes more open to seeking treatment for mental disorders, young adults are sharing their experiences on social media. This leads to a unique opportunity to see the progression of mental illness in individuals and potentially aid in diagnosis. An individual's Twitter may not be an entire representation of their overall wellbeing, but their tweets can accurately depict their distinctive traits and self-views [13]. Additionally, by using users' daily self-expressions to assess their wellbeing rather than a self-report at a specific period of time, we will be able to provide an individual with a more holistic perspective while simultaneously avoiding self-reporting bias.

We understand that social media is not fully indicative of reality, and our tool would neither serve as a replacement for therapy nor attempt to offer a technical diagnosis. Instead, we aim to analyze an individual's twitter data in order to predict their likelihood of depression and show how they present themselves on social media. We hope that this helps them keep track of their mental health and recognize that they should potentially seek treatment.

, ,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 RELATED WORK

Although depression is in itself related to the activity of certain neural circuits in the brain, there are a lot of outward behavioral and physical symptoms that come with it. One of the outward symptoms that is quite noticeable is the language an individual uses to express themselves, whether it's in speech or writing. For instance, researchers have reported that individuals with depression frequently use personal pronouns as well as "absolutist words" such as "always", "nothing," or "completely" [2]. Additionally, language usage in social media can often indicate the severity of depression that someone may have. In particular, lower depression severity was linked to a larger dispersion of negative emotions on Twitter [15].

As language usage is a prominent indicator of depression, we believe that an individual's social media activity could potentially reveal their mental state [8][7]. Media platforms such as Instagram are typically viewed as providing an overly positive representation of daily life; however, research has shown that one's Twitter usage actually provides an accurate depiction of their personality traits [13]. Gowen et al. even points out that "young adults with mental illnesses use social networking to connect with others and reduce social isolation" [9]). In fact, applications such as MoodPrism already utilize Facebook, Twitter, and music usage data to provide a user with daily feedback about their mood, mental health, and well-being [14][15]. Twitter is therefore a viable resource for detecting depression as it is already utilized by people seeking help and can provide an accurate portrayal of their personality and well-being.

The widespread use of social media creates a unique opportunity for researchers to quickly analyze large amounts of data and identify linguistic patterns that appear. By training machine learning models with these patterns, the models are able to more accurately identify mental illness and potentially improve early detection methods [16]. Twitter data, along with linguistic modeling techniques, has been used to identify patterns leading to suicidal ideation in users [6] as well as Major Depressive Disorder [12]. Short text data, such as those found on Twitter, can thus provide a great deal of insight into a user's mental state when analyzed alongside linguistic patterns found in people with depression.

Although there are challenges in obtaining adequate Twitter data, one classifier correctly identified depression-indicative posts 73% of the time [8]. Another method used a neural network model

to perform "self-harm risk classification and depression detection on social media posts" [17]. This study in particular offers insight into the performance of different models on their data set through their use of categorical cross entropy and MSE model variants. Finally, the Linguistic Inquiry and Word Count (LIWC) is a program which relies on counting words in an unsupervised manner for language analysis. LIWC has been successfully utilized in predicting depression [15] and is a comparable alternative to human labeling in predicting anxiety [10], which supports our use of a similar approach to detect depression using language analysis.

3 OBJECTIVES AND OUTCOMES

Our project deliverables include this workshop paper that discusses our motivation, related work, and creation of our machine learning models. We also designed and developed an interactive system that predicts a user's risks for depression based on their Twitter data and visualizes the results. The interactive system uses our machine learning models in the background to make predictions for risk of depression based on training data scraped from Twitter. Using the results of the model, we have included three main components of the interface: the risk percentage for depression with links to resources, a visualization of tweet attributes overtime, and a more detailed visual breakdown of the individual tweet predictions to support understanding. Lastly, we will submit a Github repository that includes our Twitter scraper, datasets, ML models, and front-end system.

4 DESCRIPTION OF WORK ACCOMPLISHED

4.1 Data Collection

In order to acquire data needed for this project, we used Twint, a publicly available data scraper compatible with Python. To determine the keywords to scrape tweets, we researched how language can be an indicator of depression. Prior studies have revealed that absolutist words and negative emotion words can be an indicator of depression [2]. Therefore, we classified tweets with those keywords as depressive. On the contrary, positive emotion words are more likely to indicate that someone is not depressed [2]. The following keywords (Table 1) are some examples of words that we scraped for, chosen by the examples of words throughout Al-Mosaiwi and Johnstone's paper.

Once all tweets were scraped, we categorized them with a binary classifier, 0 indicating non-depressive, and 1 indicating depressive. Then, we created a Python script to clean the data, in order to improve the accuracy of our model on the training data. We removed null values from our dataset and de-capitalized all of the text. Then, we removed all stop words, tags (@ symbols), hashtag symbols, emoji, URLs, and images. An important note is that we do not exclude stop words: "I," "me," "my," and "mine" because they might be used as indicators for depression. We considered spell checking the text, as well, but due to the casual nature of Twitter and commonly used abbreviations throughout social media, we concluded that spell checking our text would prevent our model from learning these shorthands.

Non-Depressive (0)	Depressive (1)
Happy, excited, good, great, cute	#feelingdown
Hope, love	#anxiety, #depressed, #depression
Playing, haha, lol, yay	Anxiety, anxious, depression, depressed
Tomorrow, weekend, morning, Friday	Disorder
Romantic, love	Antidepressant, Xanax, Prozac, medication
Nice, funny	Suicide, kms, suicidal
Better	Pain, painful, torture
Meet	Hopeless, lonely, sad, down, lost, worthless
Pretty, beautiful	Never, nothing, always, completely

Table 1: Positive and Negative Keywords used for Data Scraping

4.2 Experimental Setup/Model Details

Our experiment will give us insight into how data cleaning might affect the performance of our depression classifier. The original dataset has 1399 data points and comprises positive tweets (N=699) and depressed tweets (N=700). As a preprocessing measure, we split the data into training and test sets (80/20 split of the original data). Then we split the training set into our final training and validation sets (80/20 split of the original training set). We saved the test set for evaluation after training to check for overfitting and generalization. We packaged all of the data as a DepressionDataset class, which we defined using torch.utils.data.Dataset (used for Pytorch Dataloader module). We then used the training and validation sets for training and the test set for testing.

For our model, we decided to implement a pre-trained transformer model using the Hugging Face API. Our task falls under sequence classification, so we chose a pre-trained RoBERTa model optimized for sequence classification [1]. RoBERTa, inspired by BERT, was the optimal model choice for our work because it provides strong performance with fewer data. In comparison, other sequence classification models like BERT work slightly differently and require more data. We trained three different models roberta_v1 (trained on uncleaned data), roberta_v2 (trained on slightly cleaned data), and roberta_v3 (trained on thoroughly cleaned data).

For roberta_v1, we do not do anything before tokenizing the data. For roberta_v2, we lowercase the tweets and remove URLs, stop words, and duplicates before tokenizing the data. For roberta_v3, we lowercase every tweet and clean the dataset of stop words, emojis, URLs, duplicates, hashtags, and references to other Twitter users. We tokenize the inputs using Hugging Face's RoBERTa tokenizer to feed the model numerical representations. The input to our model is then a vector of numeric values. Our model's output is a continuous value, which we achieve by stacking a sigmoid layer on top of our logits vector, limiting our numeric domain from 0-1. Then we stack a softmax layer on top of the sigmoid layer to determine

our entire probability distribution to add up to 1. Next, we take the max probability of the vector outputted from the softmax layer and use that probability as our prediction for a person's level of depression. All output probabilities are bounded between 0 and 1, where closer to 0 indicates no depression and closer to 1 indicates depression. We decided to use both the sigmoid and softmax layers to obtain a probability score for depression, because stacking a softmax layer on top of a sigmoid layer lowers the extremes of our model's decisions, which is beneficial for user interpretability.

We use the Weights Biases logger to track our model's progress, metrics, and other analytics during training. Here is the link to our logger which contains information for all of our models: https://wandb.ai/are3010/huggingface?workspace=user-are3010.

We used the same parameters for all of our models: learning rate=5e-5, training batch size=8, and evaluation batch size=32. Roberta_v2 and roberta_v3 log model progress and performance every 15 steps and train for five epochs. roberta_v1 also logs models progress and performance every 15 steps, but only trains for three epochs. To give us a form of comparison and a better understanding of our models' performance, we included a baseline model. This baseline model decides if a user is depressed or not depressed at random. This gives our baseline model predictions by "chance" or accuracy of around 50

The metrics we used to measure our models' performance are accuracy, F1-score, and the area under the curve score for a calculated Receiver Operating Characteristic curve (ROC-AUC). We calculate these metrics for all models' performance on the held-out test set for testing and generalization purposes.

4.3 Model Results

	Accuracy	F1-score	ROC-AUC
Baseline	0.525	0.523	0.525
Roberta_v1	**0.952**	**0.962**	**0.944**
Roberta_v2	0.932	0.933	0.932
Roberta_v3	0.925	0.925	0.925

Table 2: Contains the accuracy, F1-score, and the area under the curve score for the Receiver Operating Characteristic curve for each model's performance on the held-out test set as a comparison assessment.

Our roberta models performed significantly better than the baseline, showing that our transformer model approach provides significant value to this task. On a high-level, all of the models perform very similarly. As seen in Table 2, Roberta_v1 performed with a .952 accuracy, .962 F1-score, .944 ROC-AUC, which is slightly better than roberta_v2 (.932 accuracy, .933 F1-score, .932 ROC-AUC) and roberta_v3 (.925 accuracy, 0.925 F1-score, .925 ROC-AUC). Another point is that roberta_v2 performs better than roberta_v3. These results may indicate that our model misses some important information when we clean the training data because our models' performance decreases as the data is more thoroughly cleaned. Overall, it appears that the best performing model is roberta_v1 and that a closer study of data cleaning procedures is required for this task to identify the crucial information our model needs.

4.4 Interactive Component

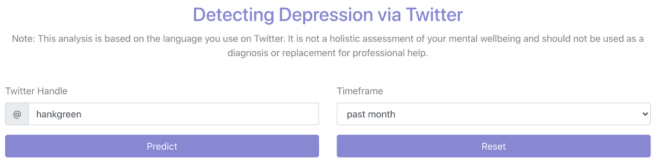

Figure 1: Frontend user inputs

4.4.1 Initial User Inputs and Form Handling. For the frontend interface, we have created a React app that accepts user inputs through a form as seen in Figure 1. To keep the system simple and reduce privacy concerns, we only ask for user input of Twitter handle and the timeframe of tweets they want to consider. We have also included a note that this analysis is only based on the language used on Twitter for more transparency and to remind users that this tool is by no means a professional diagnosis. After the user clicks the "predict" button on the form, the Flask endpoint is invoked, and a POST request is sent over to the Flask API.

Once the form data is received, the tweets are scraped from the specified Twitter account within the timeframe. If the user submits an invalid Twitter handle or the Twitter account requested did not post any tweets in the time frame selected, the POST request returns an error specifying the issue. Otherwise, the tweets are cleaned with the same script used to clean the training and testing data. We then load in the trained roberta_v3 model and tokenizer to predict the risk of depression for each individual tweet scraped. We choose to use the third model we trained, because although this model has the lowest overall scores shown in Table 2, it includes the highest level of data cleaning. We felt that this level of data cleaning was best aligned with the capabilities of what our model can interpret. For example, this version of the model removes tags from the tweets. In reality, tagging more or less people may be an indicator of depression or lack thereof, however our model is not advanced enough to pick up on these meanings and the words within usernames could cause the model to learn incorrect features.

The predictions are then averaged to output the overall risk percentage. Lastly, the user's results are sent back to the frontend through a JSON payload.

One limitation is the POST request timeout which limits the number of tweets we can scrape and predict in one request. To handle this we have a timer that will cause the Flask API to return everything it has predicted on before the fetch request timeout or error out if it could not predict anything (i.e. it spent the entire time scraping). For accounts that tweet frequently, this causes the output predictions to be based off of the most recent tweets. On average we can predict a maximum of 1400 tweets within the 2 minute Chrome POST request. This is also why we ask users to use a Chrome window.

4.4.2 Visualizations. The user receives their results with text stating their overall risk of depression (Figure 2) and two separate charts - an area chart (Figure 3) and a pie chart (Figure 4). The area chart is created with ReCharts, a library built with React and D3,

while the pie chart is created with D3. We ended up choosing to create the pie chart with D3 instead of ReCharts for more flexibility with the interactive component.

Figure 2: Overview of results and resources

Under the overall risk percentage, we have provided mental health resources for the user regardless of their predicted level of risk of depression as seen in Figure 2. Previously, we discussed only providing resources to high risk individuals, but we have chosen to provide resources to everyone as an individual's Twitter activity is not a holistic assessment of their mental wellbeing, and an individual who has a low risk according to our system may actually need these resources as well.

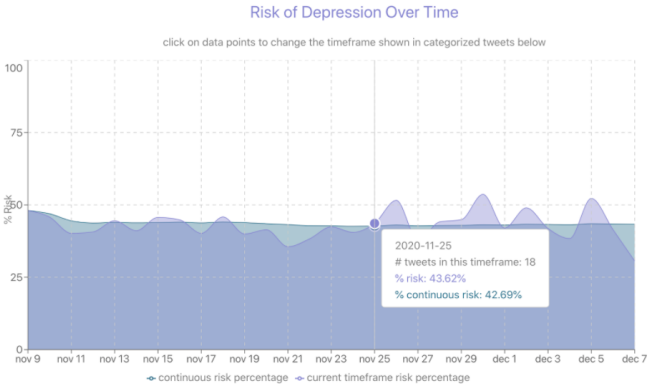

Figure 3: Visualization for overall results of user input handle and timeframe

The area chart in Figure 3 aims to allow the user to see their risk of depression throughout the entire time period they have initially selected and to see how their risk has changed throughout time. We have chosen to aggregate the data we have received initially into time frames such as days, months, seasons (3 months), or years; the type of aggregation chosen is based on the date of the first and last tweet received from the initial time frame the user selected. The main reason we have chosen to do this is to make it easier for the user to identify patterns and trends in their data that previously may not have been visible. Aggregation of the data also allows for better user experience when viewing their overall results, especially when they have posted hundreds of tweets in the time frame they have selected.

Afterwards, the area chart shows both the risk of depression for the specific time frame (ex. for a specific date or month) as well as the continuous percentage (the average percentage of depression from the beginning of time all the way up to that particular time period). Our initial prototype only displayed data points for the specific timeframe, but after iterating on the design, we decided to add visualizations for the continuous percentage as well so users can see both their risk during a specific time and their overall trends of mental wellbeing over a longer period of time. When the user hovers over a specific part of the graph, a tooltip appears displaying the dates of the time frame, the total number of tweets in that time frame, percent risk for that time frame, and percent continuous risk. Meanwhile, when the user clicks on a highlighted datapoint in the area chart, the pie chart changes its time frame to match the area chart time frame selected. This allows the user to view more details about their results in a specific time frame in hopes of providing more understandability and transparency.

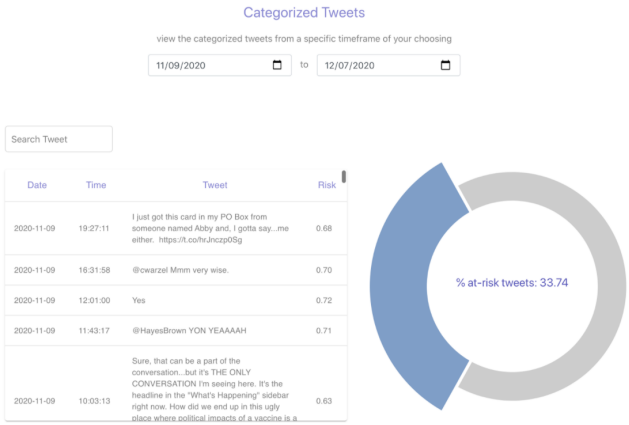

Figure 4: Detailed prediction results within selected timeframe

The pie chart, seen in Figure 4, focuses on understandability by proving a detailed look into specific dates and tweet predictions. The pie chart first splits up the data between two categories: at risk and low risk- this is based purely on whether or not the risk percentage of a specific tweet is greater or less than 0.5. We have chosen to do this so it's easier for users to find the tweets that have led to a particular risk percentage. When the user clicks on an arc in the pie chart, a table of detailed results appear. This table shows all the tweets that were placed under that particular category in the selected timeframe, as well as the time, date, and predicted percent risk of every tweet. Previously, we did not have the risk percentage of each tweet, but we chose to add it for more transparency into the exact values and confidence of each prediction. The user can also search for a particular tweet using the search bar; we have chosen to include this to assist users in exploration of a large number of tweets. The user can also change the start and end dates of the results the pie chart displays with the two date inputs above the chart itself.

5 DISCUSSION OF OUTCOMES AND IMPLICATIONS

5.1 Data Implications

We acknowledge the limitations of our dataset, and how they impact our model's accuracy. For starters, we have a small dataset. As previously mentioned, tweets are often made up of algorithms and casual conversations. The tokenization of the words in our dataset, however, may not be reflective of negative abbreviations like "kms" (kill myself), "fml" (f**k my life), and more. Similarly, our model is unable to pick up hints of sarcasm in tweets, which is commonly used throughout Twitter. Lastly, if users tweet images or websites that contain depressive or suicidal components, our scraper is unable to detect this, and our pre-processing script will remove external links. With more time and resources, we would hope to scrape external websites and photos for indicators of depression to make our results stronger and more accurate.

5.2 Model Insights and Implications

We were not able to run each model more than once because of time constraints for this project. Hence, some details on our model results may not be absolute and should be framed as speculation.

Another limitation of our model is that we are training it and testing it with a limited-in-scope dataset. It has a small data distribution and lacks diversity. We would assume that introducing more diverse data would make the task harder but help the model better learn how to classify depression. We believe that our model would benefit from a more extensive and more diverse set of training data.

Moving forward, the ground truth values associated with each example/tweet may not be entirely accurate because of how we scrape the training data from Twitter. Therefore, we hypothesize that there might be ways to explore minimizing false positives to ensure working with accurate data. This ideally will help the model perform better.

Lastly, our model surprisingly performs better when we do not clean the data. To ensure good practice, we believe it is crucial to clean the dataset. We need to identify the essential characteristics in our dataset that our model is looking for, so we do not remove them during the data cleaning process. This will allow dataset cleaning while preserving crucial information for the model.

5.3 Visualization Implications

The visualization result is highly dependent on the regular frequency a user tweets. If the user doesn't tweet regularly throughout the timeframe they have requested, the area chart isn't able to accurately reflect the way an individual's well-being changed throughout that time frame due to the gaps in data. We have tried to reduce the severity of this issue by smoothing out the area lines when we receive null values for a time frame. The results shown in the table when the user clicks the pie chart may also be confusing, as the model predicts based on the tweet that is cleaned, but the tweet we show to the user is the original tweet. Therefore, the original tweet may actually be positive due to the emojis or tags used, but we gave it a higher risk percentage due to the context of the cleaned tweet. Due to this and possibly other features of our model we run into some interesting examples. For example, a tweet that says "never been happier <3" received a prediction of at-risk for depression. This may be due to the word "never" being trained as negative or the emoji at the end. Either way, cases such as this can have an impact on the overall outcome, and this is why it is imperative that users have transparency into the predictions for each individual tweet, so that they can recognize errors in our model and adjust their trust in the results.

The visualization can also lead to interesting findings especially for those who have significant peaks or drops in the area chart. By allowing for detailed exploration of each specific tweet we allow users to understand why they received the predictions they did. The iteration from binary predictions to decimal predictions also assists users in identifying edge cases. This can also lead to users considering the reasons behind such trends and possibly assist users in considering their need for mental health help.

We use the term "risk percentage" frequently to describe our predictions, however this usage may be misleading especially in the context of depression. We chose to use "risk" rather than "likelihood" as we felt that it highlights that the predictions are not holistic, but this term still does not fully encapsulate the true meaning of our outputs. If a user has 100% risk, this does not necessarily mean that they had or have depression, but this leaves room for incorrect interpretation of our results. Similarly, if a user has a risk percentage of 0%, this does not necessarily mean that they are not at risk for depression or do not need mental health support. This case specifically can be ethically concerning. By re-evaluating the wording we choose to display in the user interface we can address these issues. One possible solution is to make it more clear that this is the predictions for the tweets themselves rather than the person behind the account. For example we could say, "@chrissyteigen's tweets showed signs of depression 47% of the time", rather than the current wording, "@chrissyteigen's risk of depression: 47%."

5.4 Future Work

Some ways to improve on our current work in the future would be to conduct user studies to better understand what type of information an individual finds helpful. Currently, our interactive visualizations are based on previous research and systems we have seen in the past, but observing individuals using our product specifically would allow us to optimize the visualization to accurately reflect the information users are looking for. We also believe that in the future, we can create results that are more detailed so that users would not only be able to see the percentage of risk they have for each tweet, but the reasoning behind it, such as the key words that led to a result. We would be interested in analyzing websites and pictures that are attached to tweets, as well, to improve our accuracy and provide users with better results. Lastly, we would like to work on increasing the scalability and performance of the system so individuals could enter a larger amount of data and receive their results in a more timely manner.

6 ETHICAL IMPLICATIONS AND RESPONSIBLE ML

Due to the complexity and severity of mental health diagnoses, we worked to create an ethical and responsible interactive system for our users. The biggest ethical challenge we faced was the potential

inaccuracy of using Twitter information to diagnose depression and the reflection of that in our system. For that reason, we include a disclaimer and additional resources on the interface, as seen in Figures 1 and 2. While our prior research has shown the success of social media as an indicator for depression, we understand that these methods can be flawed. Therefore, our interactive system will report the percentage risk someone has for being depressed, solely based on their tweets. Our system will not give a diagnosis, but rather portray potential risk based on the user's tweets.

In order to maximize the interpretability of our system, we focus on transparency and post-hoc visualizations [11]. Specifically, we clearly show the user the tweets that are depressive and non-depressive, allowing them to interact with our system and see which tweets of theirs have a higher risk value. With this, users are able to fully understand and follow their total risk percentage. We hope that through this portrayal, users can understand the outcome through their own exploration of tweets that are reflected clearly throughout the system

REFERENCES

[1] [n.d.]. Hugging Face: RoBERTa. https://huggingface.co/transformers/model_doc/roberta.html

[2] Mohammed Al-Mosaiwi and Tom Johnstone. 2018. In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science* 6, 4 (2018), 529–542.

[3] Mental Health America. [n.d.]. The State of Mental Health in America. https://www.mhanational.org/issues/state-mental-health-america

[4] American Psychological Association et al. 2019. Survey: Americans becoming more open about mental health. *Washington: American Psychological Association. Retrieved January* 2 (2019), 2020.

[5] James Butcher, Jill M Hooley, and Matthew Nock. 2019. *Abnormal Psychology.*

[6] Glen Coppersmith, Ryan Leary, Eric Whyne, and Tony Wood. 2015. Quantifying suicidal ideation via language usage on social media. In *Joint Statistics Meetings Proceedings, Statistical Computing Section, JSM.* 1–15.

[7] 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.* 47–56.

[8] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. *Icwsm* 13 (2013), 1–10.

[9] Kris Gowen, Matthew Deschaine, Darcy Gruttadara, and Dana Markey. 2012. Young adults with mental health conditions and social networking websites: Seeking tools to build community. *Psychiatric Rehabilitation Journal* 35, 3 (2012), 245.

[10] Dritjon Gruda and Souleiman Hasan. 2019. Feeling anxious? Perceiving anxiety in tweets using machine learning. *Computers in Human Behavior* 98 (2019), 245–255.

[11] Zachary C Lipton. 2018. The mythos of model interpretability. *Queue* 16, 3 (2018), 31–57.

[12] Moin Nadeem. 2016. Identifying depression on Twitter. *arXiv preprint arXiv:1607.07384* (2016).

[13] Edward Orehek and Lauren J Human. 2017. Self-expression on social media: Do tweets present accurate and positive portraits of impulsivity, self-esteem, and attachment style? *Personality and social psychology bulletin* 43, 1 (2017), 60–70.

[14] Nikki Rickard, Hussain-Abdulah Arjmand, David Bakker, and Elizabeth Seabrook. 2016. Development of a mobile phone app to support self-monitoring of emotional well-being: a mental health digital innovation. *JMIR mental health* 3, 4 (2016), e49.

[15] Elizabeth M Seabrook, Margaret L Kern, Ben D Fulcher, and Nikki S Rickard. 2018. Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. *Journal of medical Internet research* 20, 5 (2018), e168.

[16] Alina Trifan, Rui Antunes, Sérgio Matos, and Jose Luís Oliveira. 2020. Understanding Depression from Psycholinguistic Patterns in Social Media Texts. In *European Conference on Information Retrieval.* Springer, 402–409.

[17] Andrew Yates, Arman Cohan, and Nazli Goharian. 2017. Depression and self-harm risk assessment in online forums. *arXiv preprint arXiv:1709.01848* (2017).