Factors Related to Twitter Engagement in Anxiety Support Seeking

Anuraag Govindarajan, Yingchen Ma, Thomas Macheras, Elisabeth Petit - Bois

Introduction and Background

Twitter is one of the top five most popular social media websites in the US, and it is used by 22% of US adults, according to Whitney (2022). The site is an online space in which individuals come to express their perspectives and share their experiences. We are living in a time in which mental health issues are becoming less stigmatized and discussions about them are less taboo. As a result of these factors, it is inevitable that a large collection of Twitter users will use the site to share their mental health struggles and connect with others for support.

In addition, the mechanisms that users will use to connect with others will depend on the functionalities available on the website to both view others' content and post their own content. For the former, user timelines show tweets from people they follow, and the trending page shows keywords and hashtags that have been algorithmically determined to be the most talked about at any given time. For the latter, users post tweets to their followers' timelines, which may or may not include keywords or hashtags that end up on trending. Given the set of options available, those using the platform to seek a community in mental health may wonder how to maneuver their activity towards positivity and reinforcement.

For our project, we will investigate how the sentiment and the volume of the response to tweets related to a specific category of mental health, anxiety, are influenced by certain attributes, such as the presence of hashtags, profanity, positive/negative sentiment, etc. In other words, are people more likely to respond to a mental illness tweet positively or more supportively if the user includes a hashtag, includes some other textual feature, or adjusts some other controllable aspect(s) of the tweet? Answering this question could yield benefits to users who are seeking anxiety-related help, as it can give them information about how to formulate their tweet to best increase their odds of receiving a larger response and/or a more positive response.

Significance

Many people turn to social media for support regarding their struggles. This can be for a variety of reasons, such as the desire to remain anonymous and the convenience of online social interaction. Regardless of the reasoning, the number of people seeking support online is growing, and certain platforms, such as Facebook and Reddit, provide a space on their platform for those individuals to get help. However, this space is not as clearly defined on Twitter. Currently, one of the most common ways people disclose their struggles and seek support on Twitter is through the use of topically related hashtags.

Though there are hashtags associated with mental health struggles, many Twitter users still choose to disclose their difficulties without using them. Because of this, we aim to understand if there are other controllable aspects of tweets that affect engagement level. We hope to provide some guidance for individuals seeking support online, specifically regarding the levels of engagement and support which tend to result from the different methods. This would help these individuals better determine how they choose to share their struggles on Twitter.

Related Work

Social media is a powerful tool for disclosure, especially for mental health. De Choudhury (2013) examines social media as a mental illness diagnosis tool, asserting that posts expose user mental vunerabilities. The author pushes for more investigation in this research topic, urging researchers to build tools that encourage users to seek offline help. Skeels et al. (2010) tackle this challenge by designing a tool that expands support for breast cancer patients. In their design study, they ultimately build a social application to provide support beyond clinical aid. In a similar sense, where professional help cannot assist, online mental health communities (OMHCs) exist as an informal intervention tool where communities can gather, share experiences, and provide non-expert support.

Tumblr has managed to build effective social support for its transgeneder community through its hashtag feature. Hawkins and Haimson (2018) conduct user interviews that show users find comfort in their online community because of shared transition experiences. Users feel empowered and useful as active contributors on the platform. Unfortunately, these communities are subject to “hashtag hijacking” where outsiders use popular community tags to promote their content or other unrelated posts. Blair and Abdullah (2018) performed a study on Instagram where they examine hashtags related to anxiety and depression. Hijacking severely limits the amount of constructive discourse within tagged communities because off-topic posts stray from the main topic or encourage unhealthy behaviors.

On other platforms like Twitter and Reddit, users can communicate openly with one another and interact with each other using comments, likes, or upvotes. Chen and Xu (2021) examine the contagious nature of social media interaction. According to their study, positive sentiment from Reddit comments causes users to return and provide support to other users. Also, when users post positively, other users respond positively. The opposite is true as well. After conducting a study on stigmatized language on Twitter, research by Hwang and Hollingshead (2016) reasons that it is possible to combat negative spaces by educating users about mental illnesses to avoid derogatory use of language such as “crazy” or “schizo.” Their research finds that individuals who are more educated tend to avoid potentially hurtful language as opposed to those who are unaware of the impact of their statements.

When addressing how users should effectively seek mental health help online, studies often use a broad lens. Because researchers group mental health into a single search space, it is difficult to study habits and interactions characteristic of particular groups. Additionally, in their analyses, few studies examine post details beyond visible content. To address these gaps, we specifically focus on anxiety to avoid obfuscation and conflicting results from a broader topic analysis. To build a robust case for effective help-seeking, we will measure efficacy not only by post content but also by doing an in-depth analysis of the metadata, sentiment, depth, length of post replies, and other feature correlations.

Objectives, Goals, and Outcomes

Our project proposal outlined three major goals: collect an anxiety Twitter dataset, use sentiment analysis to uncover thread transformations, and examine tweet metadata to draw conclusions about support trends. Our team has managed to complete all three goals. Our anxiety Twitter dataset was curated and validated by our team using the Twitter and Reddit APIs. We used VADER Sentiment Analyzer to extract sentiment. Finally, we managed to analyze the relationship between tweet interactions and a variety of text features and metadata. Using Machine Learning models and visualizations, we were able to identify certain features of tweets that support seekers can control and likely get a larger and/or more positive response from the community.

Data

Because we were unable to find a dataset strictly featuring anxiety-related tweets, we created one using Reddit and Twitter APIs. First, we singled out five anxiety-focused subreddits to extract unigrams, bigrams, and trigrams from the "top" and "hot" categories. The five subreddits are "r/anxiety," "r/anxietyhelp," "r/healthanxiety,""r/socialanxiety," and "r/panicattack." The "hot" posts have had the most recent upvotes, while the "top" posts have the most upvotes regardless of downvotes. It was our hope that these two categories would bring out the most relatable and representative text of each subreddit, so we pulled each post title and original post text for analysis. To analyze the text and form keywords, we removed stopwords from the post title and text, found commonly occurring words and phrases, and restricted our results to items that occurred more than 10 times. Team members then selected a list of 34 promising keywords consisting of 29 bigrams and 5 trigrams.

Next, we pulled 25 tweets per keyword using the Twitter API, then proceeded to perform an annotation process to label them as anxiety-relevant (label = 1) or not (label = 0). Four team members were divided into groups of two, each of which saw both of its members annotate all tweets associated with half of the keywords. Annotators determined the relevancy of each tweet to anxiety based on simple criteria: a tweet must discuss a personal or proximal experience with anxiety and refer to anxiety as a psychiatric condition. Jokes and sarcastic tweets were allowed as long as they relate to a personal struggle. For each keyword, annotators individually determine keyword relevance as the proportion of relevant tweets in each keyword tweet set. Then, annotators paired on the same keyword evaluation compare their verdicts, and determined a keyword to be "final" if the average of the two proportions was 0.8 or greater. The average Kappa inter-annotator agreement between the groups was 0.611.

After settling on six "final" keywords, we pulled 10,000 tweets for each keyword with 5,000 tweets featuring at least one hashtag and 5,000 tweets without any hashtags. Our Twitter search was an exact keyword search on source tweets; therefore, replies, retweets, and quote tweets are not considered in the original search. We also excluded sponsored tweets and tweets with links as we found these features may derail meaningful results. After receiving all of the source tweets, we pulled all of the tweets in their reply chains for further analysis. In total, we have 42,569 source tweets and 51,414 reply tweets. 11,724 of the source tweets have replies. The date range of our data is from November 2009 to April 2022.

Approach

The goal of our project is to determine the features of a tweet, which the user has control over, that influence the volume of the response they receive and the sentiment of the response they receive. We utilize a linear regression model, trained on the features in Table 1 below, to evaluate the influence of these features on the volume of response. We quantified the volume of response in two ways. The first was with a variable we created called 'interaction_score':

(# of quotes)*10 + (# of replies)*5 + (# of retweets)*3 + (# of likes)

We assigned the weighting in the equation above based on the impact the type of interaction would have. The number of quote tweets was given the largest weight, as these provide a textual response to the tweet and move the tweet onto other Twitter users' timelines. The number of likes was given the smallest weight, as it only provides an acknowledgement/agreement of the tweet. The second way we quantified the volume of response was by the like count (# of likes), as these were the most common type of interaction. We felt it to be necessary to utilize an actual metric in addition to the variable we had created on our own.

To determine the features which influenced the sentiment of the response we utilized a linear SVM. We created two classes based on the average sentiment of the replies. If the average sentiment was positive, the tweet was labeled as being part of the positive class, and if the average sentiment was negative, the tweet was labeled as being part of the negative class. The features used in this model were the same used by the linear regression model, which can be found in Table 1.

Table 1: Extracted Tweet Features

Feature	**Explanation**
Positive Sentiment	Positivity score of tweet (using VADER)
Negative Sentiment	Negativity score of tweet (using VADER)
Contains Hashtag	0 if no hashtags present, 1 if hashtag present
Capitals	Percentage of letters which are capitalized
Mentions	0 if no users tagged, 1 if other user(s) tagged
Swear Words	Percentage of words which are swear words
Uses Profanity	0 if no profanity present, 1 if profanity present
Contains Question	0 if there is no question, 1 if there is a question
Number of Words	Normalized length of tweet
Time of Day	Normalized hour of day tweet was posted
First Person	0 if no first person pronoun(s) present, 1 if first person pronoun(s) present

In addition to these models, we also performed analysis on how the sentiment of replies on Twitter tend to evolve as their depth increases. We define the depth of a reply to be the distance from the source tweet in terms of replies. For example, if someone replies directly to the source tweet, their depth would be 1, and if someone replies to a tweet that is replying to the source tweet, their depth would be 2, and so on. We were able to calculate a depth for each reply by tracing the “referenced_tweet_id” feature for each reply up to the source tweet id.

Results

Of the two linear regression models, one predicting our ‘interaction_score’ and the other predicting the like count, we saw that the model predicting the like count performed noticeably better. The mean absolute error of the model predicting ‘interaction_score’ was 9.533, and the mean absolute error of the model predicting like count was 5.890. In Figure 1, we plotted the average prediction of our model for each true like count (only including like counts which occurred greater than 5 times). The solid red line is our model's linear regression line. Though it is not aligned with the dotted red line, which would indicate a mean absolute error of zero, it is positively correlated with the true values.

The importance of the features can be seen in Figure 2. The features which had significant positive correlation with the model were positive sentiment, negative sentiment, and mentions. Thus, it appears that more emotionally charged language, both positive and negative, tends to result in an increase in volume of response. Likewise, directly tagging another user also leads to an increase in volume of response, most likely due to these users responding. The feature which had significant negative correlation with the model was the use of swear words. Thus it appears that using more swear words in the tweet leads to other users being more hesitant to interact with the tweet.

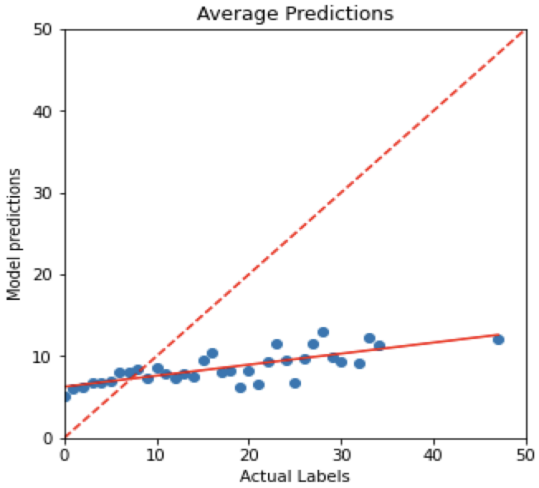

Figure 1: Average Like Count Predictions

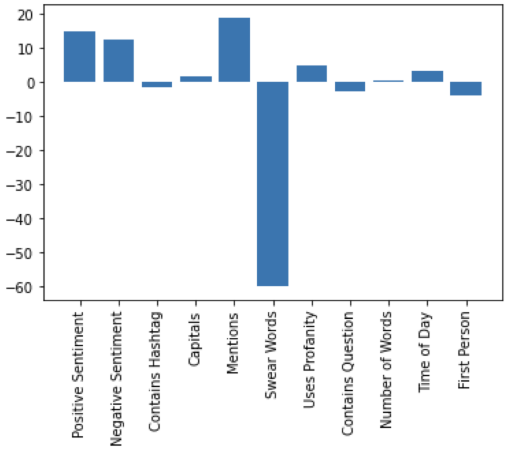

Figure 2: Feature Importance Towards Volume of Response

The performance of our linear SVM, used to evaluate the importance of the features with regards to the sentiment of the response, can be seen in Table 2. Our model's AUC is .610, indicating that it is able to differentiate between the positive and negative classes.

The importance of the features can be seen in Figure 3. The features which were correlated with the positive class were positive sentiment and containing a hashtag(s). The features which were correlated with the negative class were negative sentiment and the use of profanity. Thus, it appears that the type of sentiment used in the tweet tends to be reflected in the replies. Positive sentiment in the tweet tends to lead to positive sentiment in the replies, and negative sentiment in the tweet tends to lead to negative sentiment in the replies. Including a hashtag in the tweet could be correlated with the positive class due to it connecting the tweet to a wider, more serious topic and audience. On the other hand, it appears that using profanity within the tweet leads other users to reciprocate that negativity in the replies.

Table 2: Linear SVM Performance Metrics	
Precision	.846
Recall	.538
F1 Score	.658
Specificity	.682
AUC	.610

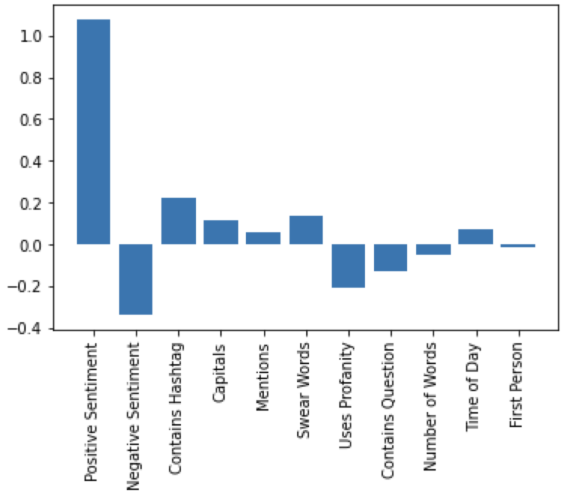

Figure 3: Feature Importance Towards Sentiment of Response

Moving on to our depth-sentiment analysis, the graph in Figure 4 is a histogram distribution of the reply depths in our reply dataset. The minimum reply depth is 1, and the maximum reply depth is 175, but as you can see in the figure, most of the reply depths are clustered between 1 and 5, meaning that conversations typically do not extend past this point.

Furthermore, we wanted to learn how the sentiment of the replies tends to change as the depth increases. Since there isn't much data available for reply depths above 20, we are only going to analyze replies from depths between 1 and 20. Figure 5 is a graph of the average overall sentiment of the tweets by reply depth. To expand on Figure 5, Figures 6 and 7 are graphs of the tweet positivity and tweet negativity by reply depth.

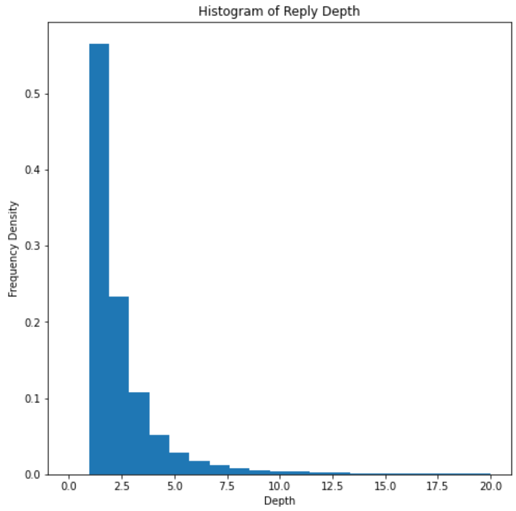

Figure 4: Distribution of Reply Depths

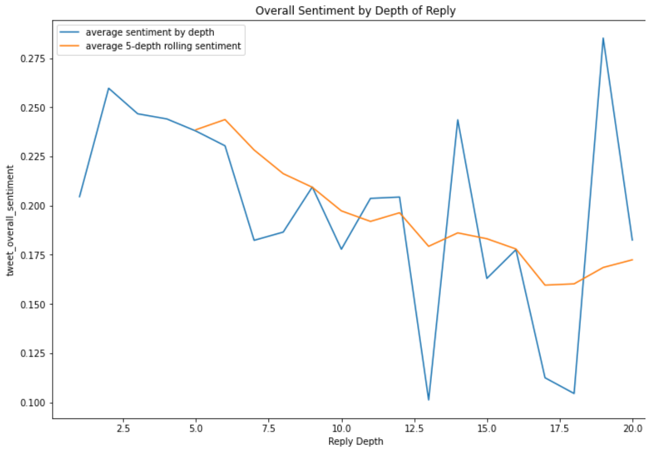

Figure 5: Overall Tweet Sentiment by Reply Depth

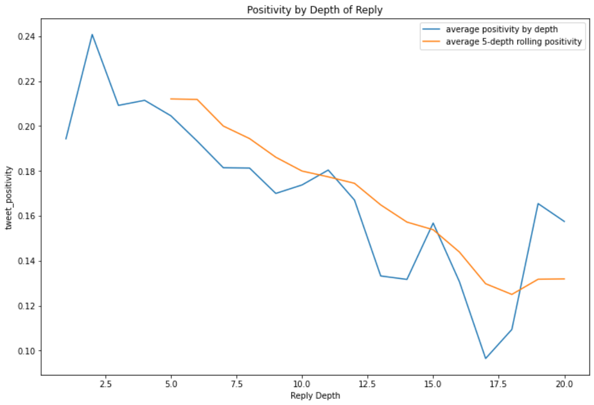

Figure 6: Tweet Positivity by Reply Depth

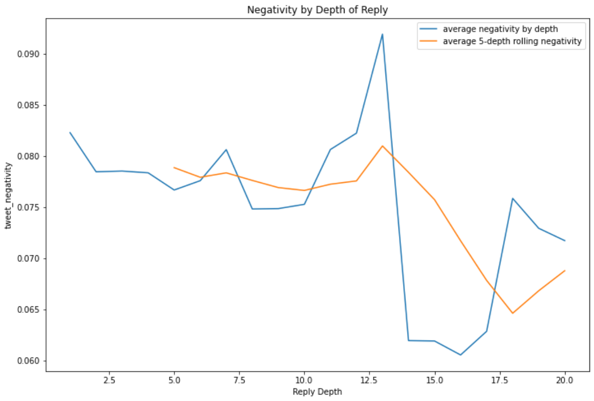

Figure 7: Tweet Negativity by Reply Depth

We can see that the rolling 5-depth overall sentiment and positivity both drop steadily from between a depth of 1 to a depth of 20, while the 5-depth negativity seems to remain constant. This might indicate that as we go deeper into the replies (up until the point where there isn't enough data) the tweet overall sentiment tends to be more negative. If we had more data from deeper in reply chains, we would be able to draw stronger conclusions about how conversations tend to evolve in anxiety-related tweet chains.

Discussion of Outcomes, Implications, and Conclusion

We have identified a few controllable aspects of tweets that can positively or negatively affect the volume and sentiment of the response they get in Twitter response threads. We determined that features such as positive sentiment, containing hashtags, and mentions can have a positive impact on volume and sentiment of engagement, and on the other hand, use of profanity and other swear words can have a strong negative impact on engagement. Outside of these features, we also learned that, surprisingly, there are many controllable aspects of tweets that were not effective predictors of engagement. These features include letter capitalization, whether or not the tweet is a question, length of the tweet, time tweeted, and perspective. Going into the project, we thought that several of these features would potentially have noticeable correlations with volume or sentiment of response, but none were according to our models.

Outside of our main models for predicting volume and engagement of responses to tweets based on various controllable aspects of the tweets, we also attempted to create models based on n-gram vectorization and narratives. Neither of these methods resulted in accurate predictions, so none of the narrative features of desire, family, money, job, or student were effective predictors of volume or sentiment of engagement. Additionally, we found preliminary evidence that sentiment of responses does tend to change slightly in the negative direction as replies get deeper into the chain. While we did not have enough data for responses at depths greater than 20, we did see a discernible pattern. This implies that few Twitter response chains, even anxiety-related, go past a certain length, and it implies that anxiety support seekers should maybe think about avoiding longer response chains.

As for future work, we can analyze trends in tweets with specific keywords. Our models described in this paper were trained and tested on a mixed dataset of various anxiety tweets. While our findings may be helpful for anxiety support seekers in general, they would benefit more from analysis that is more specific to their anxiety issues. Anxiety is an umbrella term, and analyzing keyword trends in tweets is a future step for our work.

References

Blake W. Hawkins and Oliver Haimson. 2018. Building an online community of care: Tumblr use by transgender individuals. In *Proceedings of the 4th Conference on Gender & IT* (*GenderIT '18*). Association for Computing Machinery, New York, NY, USA, 75–77. DOI:https://doi.org/10.1145/3196839.3196853

Hwang, J. D., & Hollingshead, K. (2016). Crazy mad nutters: The language of mental health. *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*. https://doi.org/10.18653/v1/w16-0306

Johnna Blair and Saeed Abdullah. 2018. Supporting Constructive Mental Health Discourse in Social Media. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare* (*PervasiveHealth '18*). Association for Computing Machinery, New York, NY, USA, 299–303. DOI:https://doi.org/10.1145/3240925.3240930

Meredith M. Skeels, Kenton T. Unruh, Christopher Powell, and Wanda Pratt. 2010. Catalyzing social support for breast cancer patients. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (*CHI '10*). Association for Computing Machinery, New York, NY, USA, 173–182. DOI:https://doi.org/10.1145/1753326.1753353

Munmun De Choudhury. 2013. Role of social media in tackling challenges in mental health. In *Proceedings of the 2nd international workshop on Socially-aware multimedia* (*SAM '13*). Association for Computing Machinery, New York, NY, USA, 49–52. DOI:https://doi.org/10.1145/2509916.2509921

Whitney, Margot. 2022. *40 Twitter Statistics Marketers Need to Know in 2022.* WordStream. https://www.wordstream.com/blog/ws/2020/04/14/twitter-statistics.

Virahonda, S. (2020, August 10). *Depressive and anxious tweets*. Kaggle. Retrieved February 14, 2022, from https://www.kaggle.com/sergiovirahonda/depression-anxiety-tweets

Yixin Chen and Yang Xu. 2021. Social Support is Contagious: Exploring the Effect of Social Support in Online Mental Health Communities. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 286, 1–6. DOI:https://doi.org/10.1145/3411763.3451644