Socially Generated Positive Reinforcement on Twitter

Jacob Johns
Department of Mathematics
Utah State University
Logan, USA
A02253549@aggies.usu.edu

***Abstract*—We explore the relationship between social positive reinforcement and posting frequency on the social media platform Twitter in a data driven manner, justified by psychological models of addiction. Our findings support the notion that there are strong relationships between users who post frequently and their responses on Twitter. The core concept is that positive reinforcement through likes, comments, and retweets will urge the user to post more frequently, much like how an addict acts with their vice of choice.**

***Index Terms*—Twitter, addiction, social media, data**

I. Introduction

Our motivation for choosing this topic came from an intersection of the authors' exposure to the Twitter API and their background in psychology. The social media industry is worth $50B and users average over 2 hours per day on various social medias and on the rise [1]. This has led to an increase seeking professional help with "social media addiction". One psychological explanation of why addictions are so perverse is due to their ability to elicit a positive reinforcement response in the user. That means that the user is stimulated in such a way that they are more likely to repeat the behavior again. Another explanation is that a craving will grow monotonically until it is indulged in. Duncan et al.,models this mathematically with a Hill function, where craving increases, but the more intense it gets, the slower the craving worsens [2]. These are the basic assumptions we make while modeling Twitter behavior with our data.

II. Related Work

A. Mathematical and Psychological Models

There has been much work done in the field of mathematical modeling of the disease of addiction. We submit to Table 1 compiled by van den Ende et al. for a more exhaustive list than offered here [3]. Dolan et al. show that social media engagement and usage are related, and can turn for the worse when a co-destruction stage is reached [4]. Alternately, they explain that when high intensity users are balanced with positively valenced engagement levels on their posts, there is a state of co-creation that is regarded as socially positive. This adds the differing perspective that excessive time on social media isn't inherently bad, especially when balanced with high engagement.

B. Data Driven Models

Our primary research question is in part answered by Di Gangi and Wasko in 2016, when they answer to what extent does user engagement affect an individual's social media usage behavior? [5] We further their work by contributing support with non-survey data, as well as more data specifically from Twitter. In their research model, frequency of use is an outcome of user experience, user engagement, and user characteristics. Our data will allow us to look closer at the user engagement; specifically the individual involvement aspect of it. The ninth hypothesis of Di Gangi and Wasko is ours as well: The higher the level of user engagement, the greater the individual usage.

III. Method

A. Data Collection

We began by requesting tweets, facilitated by Twitter's API [6]. More data was desired, but circumstances surrounding Twitter's acquisition have left API access out of budget for our research team. Due to API limits, we had to pull data in two sessions, identically so both times as shown in Fig. 1. We first got 300 random tweets, and thus 300 random users. From there, we pulled all the tweets the users made in the calendar year of 2023. Our initial dataset had over 30,000 tweets, but after cleaning and filtering, we were left with 11,000.

B. Data Cleaning

We will now explain some of the steps taken to clean and wrangle the data, and importantly why those steps are justified with this dataset and for our purposes. Tweets have a large amount of metadata and discrete data. Accordingly, we omitted many features based on redundancy (like a string version of the post id), sparse information (only 30 tweets had geolocation data), or they were not applicable to our study (the color of their profile link, among others). Furthermore, there was some data that we wanted to capture, but in a simplified form. For that reason, we saved just the presence of user mentions, rather than the username of the user mentioned in a tweet. Furthermore, to analyze the tweets of a user linearly in the time dimension, a scale for each user was created and the tweets were recorded according to when they were tweeted in relation to the tweet before it. Any outliers were also omitted, approximately 1000 observations. Finally we ensured that each

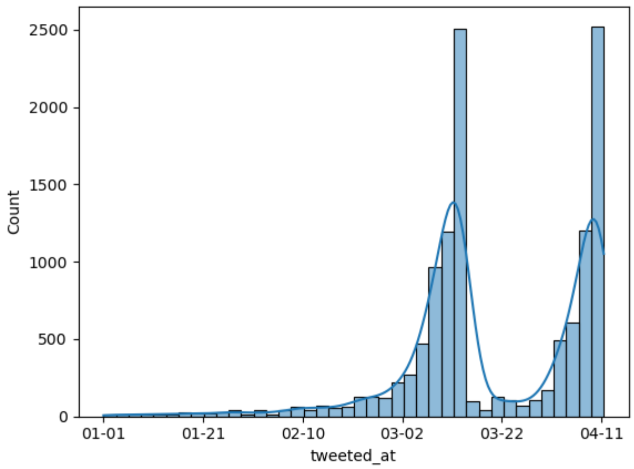

Fig. 1. Time Distribution of Tweets

remaining feature was uncorrelated with the others. This would guarantee that each one is significantly impacting our analysis. For ease of use, the data in its final form was saved as a csv file.

C. Data Summarizing and Visualization

To best understand and digest our data, we had to look at some of the general trends and summary statistics for them. These will be further explored in the Experimental Results section. Basically this entailed looking at individual outlier users, frequency plots on how often certain characteristics of tweets occur. One key point is that most features were positively skewed, as shown in Fig. 2.

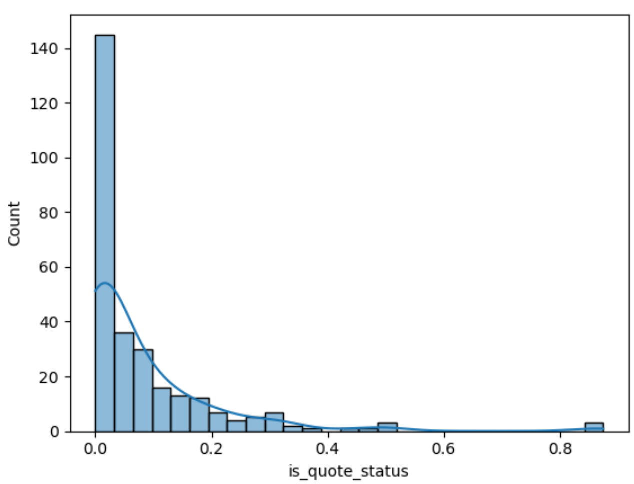

Fig. 2. Distribution of Replies in Tweets

D. Data Analysis Methods

Our analysis took one of three forms, one with regression, another statistical t-test and the other with a neural network. Regression models determine the relationship between the mean of one value and the variation in another set of variables. We will outline the regression models here, with the results being presented in the next section. In our project, univariate linear regression was used when the mean of the target variable, which was the number favorites a tweet got, is compared linearly against the variation of an independent variable, in our case, the number of retweets a tweet got. Two dimensional linear regression is similar, except the mean of the target variable is measured against a plane in 3-D space. For multidimensional regression, a hyperplane is used to measure against the mean, with the euclidean distance between the independent variables as the metric. In our multidimensional regression model, we looked at how many retweets and favorites the tweet had, along with the number of followers, friends and posts the user had. These then predicted how long the delay in time till the next tweet for the user would be. Any of these regression models can be paired with a change in basis or scale, showing an exponential relationship rather than linear.

A t-test is not a predictive model, but rather a descriptive analysis. They represent how probable it is that two samples have the same mean in their representative populations. For our analyses, we compared the number of favorites in tweets with and without media, as well as tweets that were flagged as sensitive or not. This allows us to consider the variations and distribution of those tweets into account while still being able to justly compare them.

Our final predictive analysis technique used was neural networks. These are models in which the features are weighted and scaled and then related one to another in layers. The model can then predict which value a novel datum would produce by traversing the weighted network.

IV. Experimental Results

A. Descriptive Models

We begin by discussing analyses that are insightful and interesting, but lack connection to our research hypothesis. One metric captured by the Twitter API is if a tweet is potentially sensitive or not. It is important to note that it makes no distinction between if the tweet was flagged as such by the author of the tweet or by Twitter itself. As previously explained, we performed a t-test on this data and found that sensitive tweets received statistically significantly less favorites than those tweets which were not flagged as sensitive. One explanation for this is that some sensitive content isn't available to minors or users without a birthday on their profile [7].

TABLE I
T-test Results for Safe and Sensitive Tweets

t-statistic	p value	Sensitive Mean Favorites	Safe Mean Favorites
-11.49	3.00e-18	0.50	2.53

A second t-test was performed to describe the differences in tweets with and without media. We used the metric of the

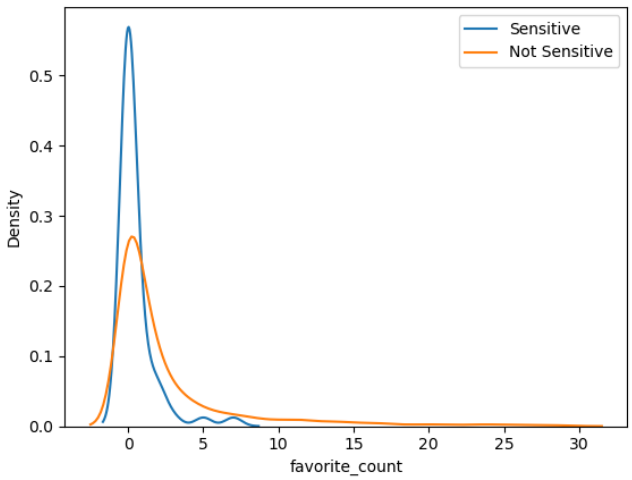

Fig. 3. Distinct Distributions of Safe and Sensitive Tweets

number of favorites a tweet received to differentiate between the two populations. Again, we found that tweets with media received statistically significantly less favorites than those without media. This defies expectations, as media is typically regarded as engaging and more likely to evoke a response, but such was not the case.

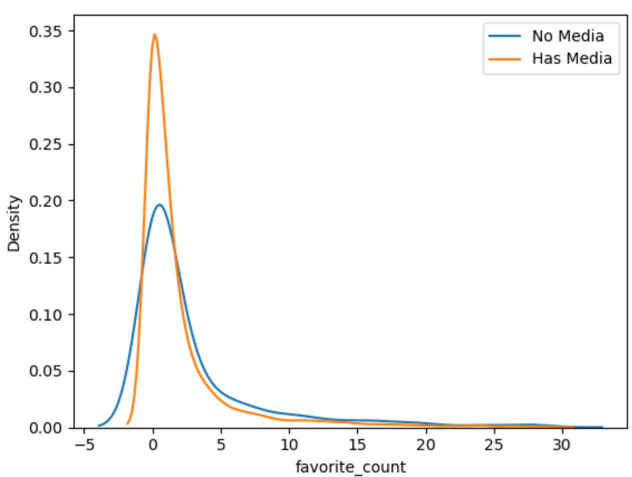

Fig. 4. Distinct Distributions of Tweets with and without Media

TABLE II
T-TEST RESULTS FOR SAFE AND SENSITIVE TWEETS

t-statistic	p value	Media Mean Favorites	No Media Mean Favorites
4.69	2.99e-6	2.84	2.01

B. Predictive Models

We now will discuss some predictive models that are theoretically adjacent to our research and warrant an in depth study. First of which was a linear regression correlating the number of retweets a tweet got with its number of favorites. We assume that retweets only come from users who also favorite the tweet as well. Doing so validates our assumption that they're directly linearly correlated. We found that there was some positive correlation, but there was little confidence in it. As such, our findings are not presented in absolute terms.

$$\text{Favorites} \approx 5.5 \times \text{Retweets} + 2$$

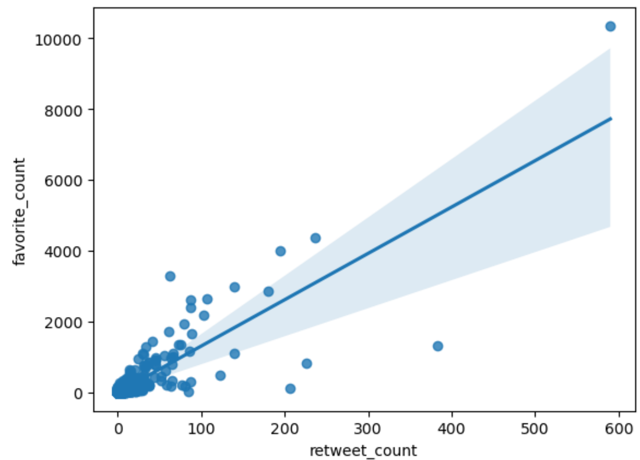

Fig. 5. Correlation between Retweets and Favorites

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         favorite_count   R-squared:                       0.725
Model:                            OLS   Adj. R-squared:                  0.725
Method:                 Least Squares   F-statistic:                 2.862e+04
Date:                Tue, 02 May 2023   Prob (F-statistic):               0.00
Time:                        13:49:02   Log-Likelihood:                -62188.
No. Observations:               10881   AIC:                         1.244e+05
Df Residuals:                   10879   BIC:                         1.244e+05
Df Model:                           1
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
const            -0.9811      0.708     -1.386      0.166      -2.368       0.406
retweet_count    13.2593      0.078    169.166      0.000      13.106      13.413
==============================================================================
Omnibus:                    16077.955   Durbin-Watson:                   2.100
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        538913718.191
Skew:                          -7.455   Prob(JB):                         0.00
Kurtosis:                    1093.160   Cond. No.                         9.08
==============================================================================
```

Fig. 6. Statistical Results for Univariate Linear Regression

To hopefully add to our confidence in predicting the number of likes a tweet received, we then considered a two dimensional linear regression. Here we looked at the number of retweets a tweet got along with the number of followers the user had. We assume that the number of favorites someone gets is affected by how many people retweet it and how many followers the user has because of our previous analysis as well as the fact that those with more followers have more exposure. By doing so, we were able to marginally raise our confidence in the model. However, it is interesting to note that in the two dimensional case the coefficient for the number of followers is negative, meaning that the more followers a user has, the less favorites our model predicts the tweet will get.

This is opposite to the one dimensional case, where favorites and number of followers is positively correlated. We offer no satisfactory explanation as to why this phenomenon occurs.

$$\text{Favorites} \approx 5.36 \times \text{Retweets} - 0.00002 \times \text{Followers} + 2$$

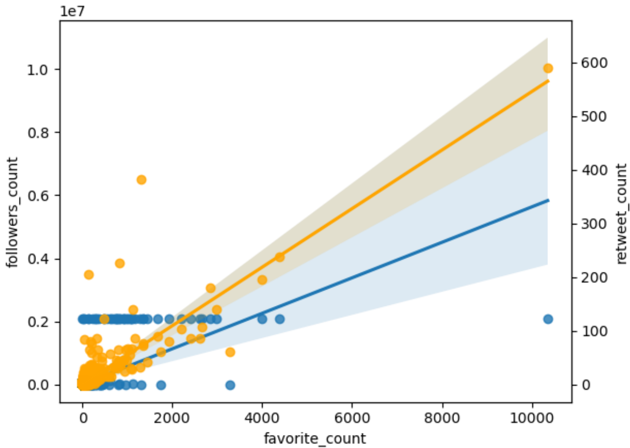

Fig. 7. Correlation between Retweets and Followers with Favorites

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         favorite_count   R-squared:                       0.777
Model:                            OLS   Adj. R-squared:                  0.777
Method:                 Least Squares   F-statistic:                 1.899e+04
Date:                Tue, 02 May 2023   Prob (F-statistic):               0.00
Time:                        13:49:03   Log-Likelihood:                -61030.
No. Observations:               10881   AIC:                         1.221e+05
Df Residuals:                   10878   BIC:                         1.221e+05
Df Model:                           2
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const              -2.7654      0.637     -4.340      0.000      -4.014      -1.516
followers_count     0.0003   5.34e-06     50.796      0.000       0.000       0.000
retweet_count      11.4870      0.079    146.080      0.000      11.333      11.641
==============================================================================
Omnibus:                     7844.630   Durbin-Watson:                   2.104
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        661837676.769
Skew:                           1.347   Prob(JB):                         0.00
Kurtosis:                    1211.220   Cond. No.                     1.34e+05
==============================================================================
```

Fig. 8. Statistical Results for Bivariate Linear Regression

In an act of hubris, we tried to extend this model to be multidimensional, with five features explaining the time till the next tweet by the user. We considered how many retweets and favorites the tweet had, along with the number of followers, friends and posts the user had to predict the time till the next tweet. Unfortunately, such an analysis was unproductive as there was too much variance in each of the features and too little correlation between them to model the time delay properly. We failed to find support for our 5 dimensional linear regression.

Upon analyzing the relationship between retweets and the time differential, a supposed exponential relationship is seen. However, this exponential decay is not a good fit or model for prediction as when the data are transformed into a log-log scale, there is no discernible pattern. We performed this

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                minDiff   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     3.923
Date:                Tue, 02 May 2023   Prob (F-statistic):            0.00149
Time:                        13:49:04   Log-Likelihood:            -1.0246e+05
No. Observations:               10881   AIC:                         2.049e+05
Df Residuals:                   10875   BIC:                         2.050e+05
Df Model:                           5
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const             751.6065     31.422     23.920      0.000     690.013     813.200
retweet_count       5.0909      6.099      0.835      0.404      -6.865      17.047
favorite_count     -0.3752      0.432     -0.868      0.386      -1.223       0.472
followers_count -8.036e-05      0.000     -0.299      0.765      -0.001       0.000
friends_count      -0.0180      0.010     -1.768      0.077      -0.038       0.002
statuses_count     -0.0013      0.000     -3.143      0.002      -0.002      -0.001
==============================================================================
Omnibus:                    19879.978   Durbin-Watson:                   1.995
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         35474083.042
Skew:                          13.465   Prob(JB):                         0.00
Kurtosis:                     281.423   Cond. No.                     1.47e+05
==============================================================================
```

Fig. 9. Statistical Results for Multivariate Linear Regression

analysis on those tweets which were above a two sigma threshold for the number of retweets got on a user by user basis. We repeated this attempt at exponential decay for both the number of retweets and the number of favorites a tweet got. Neither were conclusive. Highlighted in Fig. 11 and in Fig. 13 we see that there is no pattern to the distribution of time among retweets. Nor is there any rhyme or reason to how the z scores of the tweet in their respective feature are distributed. That is to say, a tweet that caused a quick turnaround in the user using Twitter again is no more likely to be above or below average. Thus we found no support for our hypothesis that abnormally high levels of retweets or favorites on a tweet would cause the user to post more frequently.

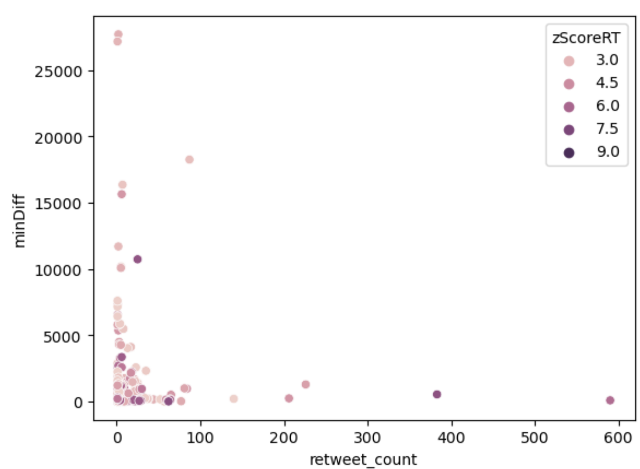

Fig. 10. Abnormally high retweets

We then considered the opposite: What is about if abnormally low levels of social stimulus decreased the likelihood of posting? In physiological terms, this would be as if responses on a tweet were a negative punishment— the withholding of social stimulus triggers the user to decrease the behavior of posting on Twitter. Accordingly, we analyzed in the same

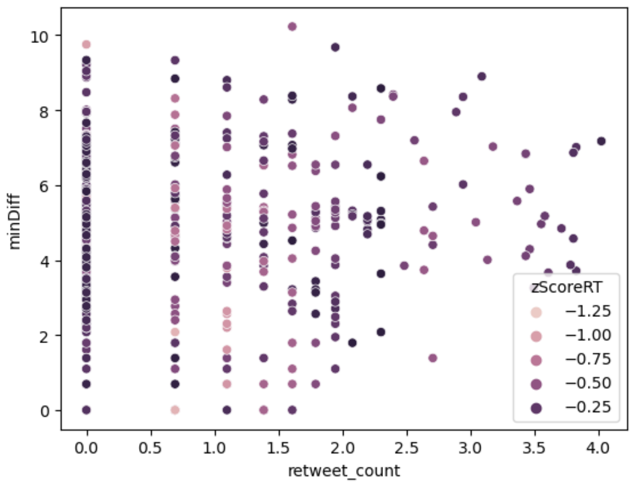

Fig. 11. Abnormally high retweets - log scale

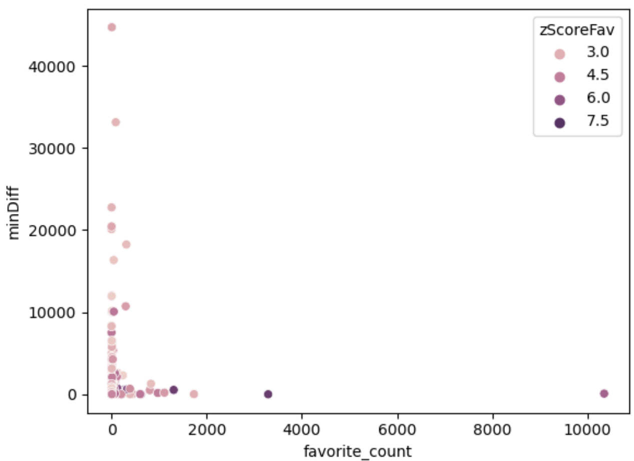

Fig. 12. Abnormally high favorites

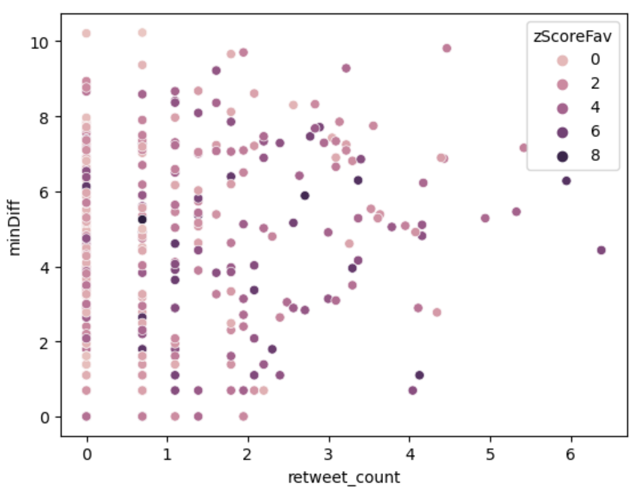

Fig. 13. Abnormally high favorites - log scale

manner previously, finding a supposed exponential decay relationship. This was found both in tweets with a below average number of favorites and below average number of retweets. However, much like before, once the data were transformed, there was no familiar distribution or pattern, meaning that an exponential model would not fit it well either. Thus we found no support for our hypothesis that abnormally low levels of retweets or favorites on a tweet would cause the user to post less frequently.

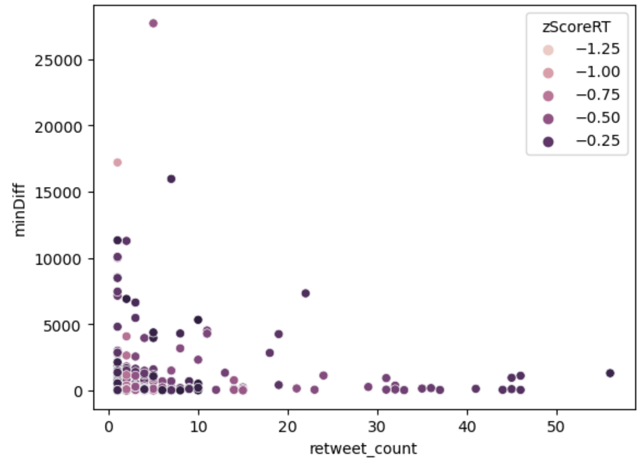

Fig. 14. Abnormally low retweets

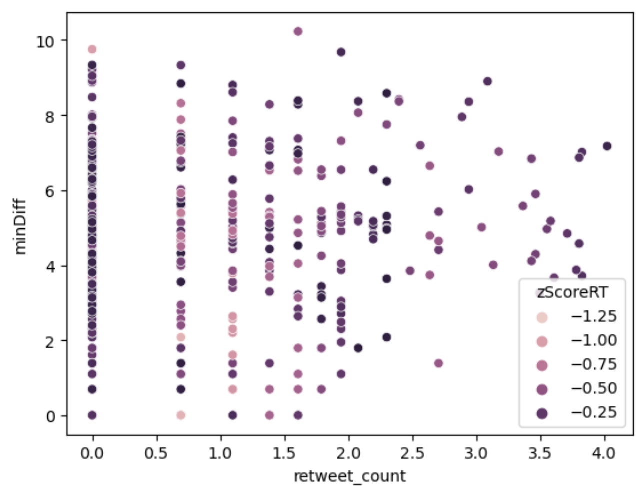

Fig. 15. Abnormally low retweets - log scale

For our final analysis, we looked at how a neural network would be able to detect patterns unseen to us in the data. After a few permutations, it was found that a neural network with 4 sublayers of size 10, 14, 5, and 3 fit the data best. However, such a predictive model was unsatisfactory. There was no confidence in it, and the range of error was astounding. We can attribute this again to the large variance in the time when people tweet. There is just too much variance in when people tweet to accurately describe it with just a few features and no considerations for individual behavior.

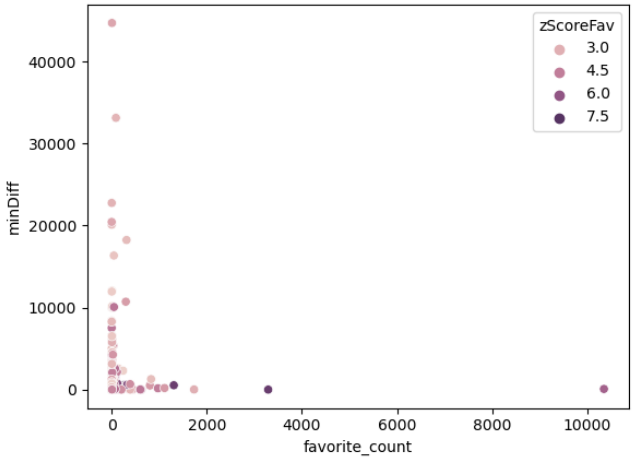

Fig. 16. Abnormally low favorites

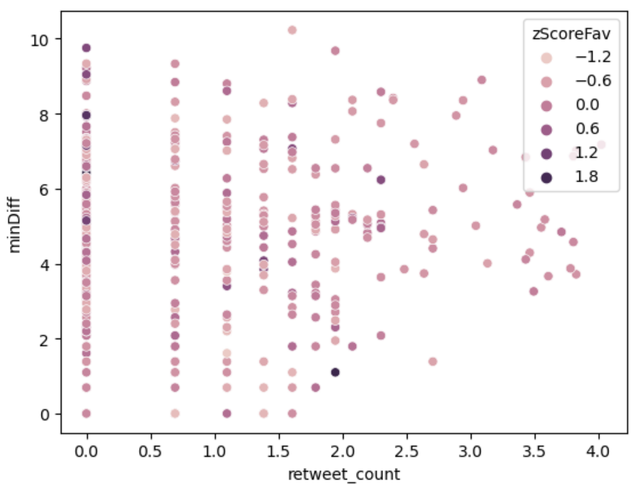

Fig. 17. Abnormally low favorites - log scale

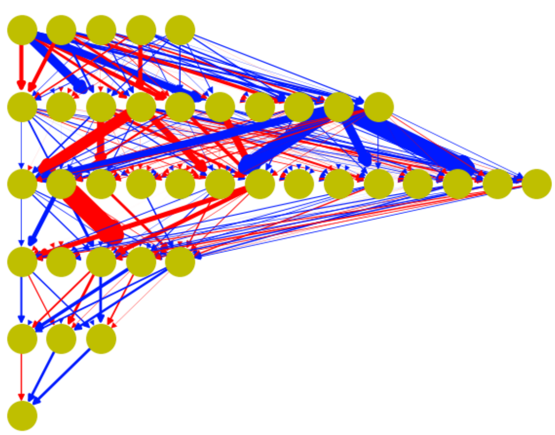

Fig. 18. Graphical Representation of Neural Network

TABLE III
PERFORMANCE RESULTS FOR NEURAL NETWORK

R^2	MSE
0.016	6.9e6

V. CONCLUSION

In conclusion, modeling human behavior is hard to do, especially with limited features of their behavior. While the psychological principles still apply, clearly there is more than meets the eye when they are applied in real life. We failed to find enough evidence to reject the null hypothesis. This means that there is not sufficient evidence to indicate that an above average number of likes or retweets can predict with confidence any timeline on when the user will post again. Furthermore, we failed to find evidence to the negative of our hypothesis. That means that there is not sufficient evidence to indicate that a below average number of likes or retweets can predict with confidence when the user will post again. Future work would entail looking at the content of the tweets as well and following users across a set amount of time. Further data would be needed, especially data that the Twitter API can't offer. Alternative approaches, such as time series analysis, transformers, Long-term Short-term models or rule discovery. Finally, we report that despite the hardships in analyzing this data, it has been productive as a learning experience on how to collect, preprocess, visualize, summarize and analyze data.

REFERENCES

[1] https://datareportal.com/reports/digital-2022-global-overview-report
[2] https://epubs.siam.org/doi/epdf/10.1137/18M121410X
[3] https://www.sciencedirect.com/science/article/pii/S0306460321003865#b0140
[4] https://www.tandfonline.com/doi/full/10.1080/0965254X.2015.1095222
[5] https://www.researchgate.net/profile/Paul-Di-Gangi-2/publication/299357565_Social_Media_Engagement_Theory/links/5702df4808aea09bb1a305ab/Social-Media-Engagement-Theory.pdf
[6] https://developer.twitter.com/en/docs/twitter-api
[7] https://help.twitter.com/en/rules-and-policies/media-policy