

FROM MUSLIM TO MINDFULNESS: TWITTER RETWEETS BASED ON SENTIMENT AND LINGUISTIC ANALYSIS OF SELECTED CURRENT KEYWORDS

Alan R. Peslak, Penn State University, arp14@psu.edu

ABSTRACT

Twitter has become one of the most immediate ways to record public sentiment about issues, topics, and personality. The immediacy and ubiquitous nature of tweets has even risen to be the most popular means of communication for the President of the United States. Our study examines Twitter keywords on one particular day in 2017 in attempt to analyze whether there are certain characteristics of tweets that make them popular. We measure popularity by number of retweets and we analyze nearly 100 variables that may affect number of retweets. Our analysis of controversial and non-controversial keywords discovers some significant findings affecting retweet popularity.

Keywords: Social networking, Twitter, Linguistic analysis, LIWC, Sentiment analysis, Content analysis, Qualitative analysis

INTRODUCTION

This study is an exploration of Twitter sentiment and linguistics to determine what the current state of popular discourse is in the United States. Twitter is the most popular form of microblogging or posting of short messages on the Internet. These messages are publicly available to all who “follow” the specific communicator. The short messages are limited to 140 characters per message. To receive the message the “follow” process requires an individual to subscribe to Twitter service and then select the individuals or organizations you wish to follow by finding their Twitter handle (name of account) and adding them to a your list of “following”. You will then receive all messages from that user that they post to their account. These short message posts are called tweets. You can follow any number of users and any of number of users can follow you if you post messages. There is no acceptance by the other party on either side. You can follow or people can follow you. There is a mechanism for blocking specific users from following you, however, this is an opt-out feature.

Twitter has become a very popular Internet communications tool since its launch in July of 2006. Growing from the first tweet in March of 2006, it now has grown to over 500 million tweets or 140 messages each day. A day’s worth of tweets would fill a 10 million page book. There have been over 1.3 billion accounts created over the past 10 years and there are 310 million active users. It is also one of the most popular forms of social media with 29.2% of social media users being Twitter users. Surprisingly, 83% of world leaders are on Twitter, culminating in the President of the United States currently using Twitter as his most active source of communications. (Smith, 2016).

One of the major sources of Twitter communications is the ability to resend or retweet (RT) an already posted original message. A popular message from a user can be resent from that user’s account and is classified as a retweet. The number of retweets thus indicates popularity of a message such that users want to repost that message to their followers. The most popular retweeted message had 3.3 million retweets. (Smith, 2016). With retweets being a major source of content in the Twittersphere and a readily available variable which can indicate popularity of the message, this manuscript attempts to review variables that may affect the number of retweets of a tweet and hence its popularity. The methods used to explore potential influencing variables were specifically available descriptive variables related to the tweet and available from Twitter, as well as a comprehensive sentiment and linguistic analysis of specific words or terms (keywords) that are included in the tweet. The keywords are topic words. Our analysis includes analysis of which variables affect popularity as measured by retweets identified with that keyword. We hypothesize that popularity of tweets as measured by retweets will be affected by certain variables. Our study is an attempt to determine those variables and whether they vary by topic.

LITERATURE REVIEW

The use of sentiment analysis and linguistic analysis has been studied by the author in several manuscripts. (Peslak, 2016a, Peslak, 2016b). As discussed in previous publications, sentiment analysis and linguistic analysis are common methods of research in communication analysis. The usage of linguistic analysis and specifically the use of LIWC (Linguistic and Word Count) software for research purposes has also been extensive. Robinson, Navea, and Ickes (2013) used LIWC analysis of students written self-introductions to grades that students achieved. Bell, McCarthy, and McNamara (2012) used LIWC to review whether there were differences between males and females in linguistic styles. Sexton and Helmreich (2000) used LIWC to study airline cockpit communications to determine see if there were errors in judgment or performance of duties. . Back, Kufner, and Egloff (2011) studied the New York City Twin Tower 9/11 communications using LIWC. There are many more examples of the use of LIWC used for scholarly research.

Linguistic and Word Count (LIWC) software (Pennebaker, Booth, Boyd, and Francis, 2015) is one of the most accepted and popular linguistic analysis tool. “The way that the Linguistic Inquiry and Word Count (LIWC) program works is fairly simple. Basically, it reads a given text and counts the percentage of words that reflect different emotions, thinking styles, social concerns, and even parts of speech. Because LIWC was developed by researchers with interests in social, clinical, health, and cognitive psychology, the language categories were created to capture people’s social and psychological states. The text analysis module then compares each word in the text against a user-defined dictionary. As described below, the dictionary identifies which words are associated with which psychologically-relevant categories.

After the processing module has read and accounted for all words in a given text, it calculates the percentage of total words that match each of the dictionary categories. For example, if LIWC analyzed a single speech that was 2,000 words and compared them to the built-in LIWC2015 dictionary, it might find that there were 150 pronouns and 84 positive emotion words used. It would convert these numbers to percentages, 7.5% pronouns and 4.2% positive emotion words.”(Pennebaker Conglomerates, 2015).

Linguistic evaluation using LIWC on Twitter has been achieved within the past by numerous researchers. Tumasjan, A., Sprenger, T. O., Sandner & Welpe (2010) used LIWC to observe election predictions from Twitter feeds. González-Ibáñez, Muresan, & Wacholder (2011) tested sarcasm in Twitter messages using LIWC evaluation. Stieglitz, S., & Dang-Xuan, L. (2012) suggest that positive and negative emotions and affect, significantly correlate with number of retweets. Coppersmith, Harman, & Dredze, M. (2014) used LIWC evaluation of Twitter feeds to predict mental health.

Bing Liu (2012) defines. “Sentiment analysis, also called opinion mining, is the field of study that analyzes people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes.” Sentiment Analysis is the review of written or other forms of communication or qualitative data to determine a quantifiable and comparable measure of some form of feeling in the communication or data.

Pak & Paroubek, P. (2010, May) provided the seminal study on the use of Twitter for Sentiment and Opinion Mining. Their study developed a method for developing a sentiment classifier from Twitter data. Pang and Lee (2008) suggest that one of the most studied areas of sentiment analysis is sentiment valence which is the degree of positive or negative sentiment. A simple concept is to determine whether a particular communication is positive or negative and the degree to which it is either. Eguchi and Lavrenko (2006) suggest. “One of the first and still most used method of sentiment analysis is keyword analysis, where a text is reviewed word by word and compared against a dictionary. This dictionary has been previously prepared and will classify each word in its dictionary into a sentiment rating.”

METHODOLOGY

In order to sample this Twitter discourse, a day was randomly selected in 2017, March 15, 2017 and the front page of the New York Times was reviewed to find key news items and terms that would serve as keywords which would be used to mine Tweets based on these keywords. A selection was made of both political and religious terms (hence

controversial) as well as non-controversial terms. From this selection, two controversial terms were analyzed as well as two non-controversial terms. The terms selected were:

Controversial/Political/Religious: Donald Trump and Muslim

Non-controversial: Snowstorm and Mindfulness

Donald Trump and his presidency has generally represented polarizing views across the US and extensive divided attitudes have been expressed about the Muslim religion and immigration. Mindfulness is a meditative state and is generally not viewed with controversy. Also March 15 was the day after a large East coast snowstorm and significant press was devoted to this non-controversial weather event. The selection of these terms was the decision of the authors but all were the subject of front page stories in the New York Times on March 15, 2017.

Once these keywords were selected, Twitter posts were extracted that related to these keywords. The process used to capture these Twitter posts was to extract 1000 Twitter posts for each term using BirdIQ from North Concepts. BirdIQ is a .jar executable which will search for current tweet based on an input keyword. It will then return an excel spreadsheet that includes all available fields for each tweet from the Twitter datastream for each tweet. It also returns summary sheet for hashtag entries, mentions, urls, users, retweets, favorites, and words. The Retweet sheet was used for our analysis since it contains the variables: Retweet Count, Full Text of Tweet, User Followers Count, User Following Count, User Tweets, User Favorites Count, and keyword. As noted, we used Retweet Count as our dependent variable to determine popularity of tweets. The other tweet specific descriptives were able to be explored as independent variables. Added to the spreadsheet was a sentiment variable calculated by the publicly available sentiment calculator from Microsoft Azure Machine Learning. This variable develops an overall measure of sentiment ranging from 0 (negative) to 1 (positive) with .5 being neutral. A specific multi-decimal rating is specifically developed e.g. .54678 from this Microsoft Excel plug-in.

We then imported the Retweeted text into LIWC (Linguistic and Word Count). LIWC software results produce 93 unique measures from each of its linguistic analyses. These measures range from parts of speech to emotional categories to word counts. For the most part these are expressed by a percentage of total words mapping to the dictionary category of each measure. The exceptions are several relating to word counts as well as calculated emotional measures. Appendix A has a complete list of the BirdIQ independent variables as well as the 93 LIWC variables. There is also a definition of each or examples of words that meet the LIWC category.

With RetweetCount as the dependent variable, a regression analysis was run on each of the four keyword results using all 99 LIWC and BirdIQ values as independent variables. IBM SPSS 23.0 was used to develop the regression analysis and the results follow.

RESULTS

The four keywords used were Trump, Muslim, Mindfulness, and Snowstorm. Again the first two were analyzed due to their current controversial attitudes. They were selected to see if there were common variables that predicted tweet popularity for controversial issues. Mindfulness and snowstorm were selected as non-controversial issues. The objective here was similarly to see if there were significant variables affecting popularity of tweets on non-controversial issues.

The overall results of the regression analysis found three of the four regression results to be significant at $p < .001$. This suggests that there are sentiment and linguistic variables that did influence popularity of specific tweets based on number of retweets. However, for each keyword there was limited consistency on which variables did influence retweets. Each keyword is now analyzed. The first data analyzed were the tweets based on Donald Trump keyword. Of the 1000 selected tweets on Trump on 3-15-2017, 318 had at least one retweet. These tweets were analyzed by BirdIQ, Microsoft Azure Machine Learning Sentiment Analysis, and LIWC. The data collected from this analysis were then examined via linear regression with Retweet count as dependent variable and the 99 variables as independent variables. The regression results had an r square of .636 and was significant at $p < .001$. Of course, many of the variables were not found to be significant, but 20 variables were significant at $p < .10$ and are shown in table 1.

Table 1. Trump regression analysis

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	1343.187	2833.424		0.474	0.636
Dic	501.942	128.048	1.163	3.92	0
Affiliation	2383.163	521.394	0.725	4.571	0
Home	4651.451	643.694	0.429	7.226	0
Paranth	-975.082	298.004	-0.162	-3.272	0.001
Function	-798.157	287.722	-1.294	-2.774	0.006
Leisure	-641.123	235.697	-0.153	-2.72	0.007
You	1524.694	592.885	0.295	2.572	0.011
Adj	507.477	196.984	0.208	2.576	0.011
Money	-510.456	216.121	-0.162	-2.362	0.019
Family	-1595.31	760.965	-0.216	-2.096	0.037
I	896.472	435.106	0.216	2.06	0.041
Anx	-954.923	488.916	-0.121	-1.953	0.052
Negate	743.434	381.88	0.164	1.947	0.053
Apostro	-281.701	147.574	-0.108	-1.909	0.058
Sexual	1791.572	959.877	0.329	1.866	0.063
See	429.787	230.968	0.108	1.861	0.064
Quant	-483.932	260.77	-0.123	-1.856	0.065
Assent	1433.019	789.807	0.114	1.814	0.071
Sixltr	105.95	60.775	0.12	1.743	0.083
Insight	-593.776	359.948	-0.157	-1.65	0.1

A detailed analysis of each of the 20 variables is impractical for this manuscript, we therefore focus on general themes revealed as well as those with the highest impact as measured by their Beta or Standardized Coefficient. In general, retweet popularity was influenced by many tangible and basic needs such as money, family, sex, home, leisure and affiliation. But the two variables that had the largest impact were Dictionary words and Function words. Dictionary words were positively correlated to retweets. Dictionary words are percent of words in the tweet that are in the LIWC dictionary. Complex and technical words are less likely to be in the dictionary, thus a high score here suggests that simple, easy to understand, and tweets using common words are more likely to be retweeted than more difficult words. Function words on the other hand are negatively correlated with retweets. Function words are non-content words such as pronouns, articles, prepositions, conjunctions, quantifiers, and common adverbs. They do not include any nouns or verbs or real action but serve generally as connectors and fillers. The fact that function words are negatively correlated means that retweets count favors tweets that in include direct and simple content, not function or filler words. Trump retweets are most affected by action and simplicity.

Table 2. Muslim regression data

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
I	2040.428	415.785	0.779	4.907	0
Cogproc	-1514.09	369.44	-1.23	-4.098	0
Insight	2143.262	447.033	0.668	4.794	0
Discrep	1572.439	381.563	0.411	4.121	0
Certain	1837.3	448.329	0.481	4.098	0
Differ	1967.195	452.552	0.77	4.347	0
Tentat	1136.456	373.982	0.344	3.039	0.003
Prep	848.369	292.132	0.758	2.904	0.004
Negate	1223.121	420.56	0.394	2.908	0.004
Cause	1220.407	423.821	0.455	2.88	0.005
Authentic	-135.236	49.799	-0.399	-2.716	0.007
Ipron	862.928	315.658	0.428	2.734	0.007
Function	-740.635	274.411	-1.66	-2.699	0.008
Article	785.941	313.135	0.514	2.51	0.013
Family	-890.016	375.955	-0.213	-2.367	0.019
Adverb	639.944	275.66	0.248	2.321	0.022
They	878.96	405.336	0.237	2.168	0.032
Sexual	2229.433	1035.959	0.266	2.152	0.033
Tone	-66.135	30.975	-0.323	-2.135	0.034
Anger	-938.728	460.366	-0.324	-2.039	0.043
You	886.205	450.22	0.282	1.968	0.051
Auxverb	742.134	377.29	0.524	1.967	0.051
Clout	128.563	67.583	0.453	1.902	0.059
Affiliation	882.249	464.729	0.366	1.898	0.06
Conj	530.152	282.991	0.264	1.873	0.063
Social	-372.194	204.773	-0.382	-1.818	0.071
SentN	5544.609	3237.261	0.15	1.713	0.089

The next data analyzed were the tweets based on Muslim keyword. Of the 1000 selected tweets on Muslim on 3-15-2017, 248 had at least one retweet. These tweets were analyzed by BirdIQ, Microsoft Azure Machine Learning Sentiment Analysis, and LIWC. The data collected from this analysis were then examined via linear regression with Retweet count as dependent variable and the 99 variables as independent variables. The regression results had an r square of .567 and was significant at $p < .001$. Of course, many of the variables were not found to be significant, but 27 variables were significant at $p < .10$ and are shown in table 1. A detailed analysis of each of the 27 variables is impractical for this manuscript, we therefore focus on general themes revealed as well as those with the highest impact as measured by their Beta or Standardized Coefficient.

In general, retweet popularity was influenced by somewhat different factors than the Trump tweets. Here we see anger, tone, clout, certainty, cause, negation as significant factors. This suggests a more emotional affect to retweeting texts with keyword Muslim. The highest impact variables were I, differ, prep, cognitive process, and insight. The high use of I in Muslim retweets suggests a much more personal aspect to the religious related retweets. The cognitive processes

and insight variables suggest a more thoughtful and reasoning approach. In addition, the differ variable relates to the presence of qualifiers such as but and else suggesting multiple sides of the issue. In general, it seems that the Trump retweets were more simplistic and direct but the Muslim retweets were more thoughtful and personal.

There were areas of commonality for the controversial Trump and Muslim tweets. The following variables were significant for both keywords: affiliation, family, function words, I, insight, negate, sexual, and you. Many of these variables suggest personal attitudes that may appeal on an emotional level with individuals. Both were influenced by belonging (family, affiliation) and both were influenced by personal pronouns (I and you). Both also had higher retweets for simpler tweets (function words) but both also were affected by insight and sex. One major difference here though was that Trump retweets had lower insight and Muslim retweets has higher.

The next data analyzed were the tweets based on non-controversial keyword topics. Mindfulness and snowstorm were selected from the New York Times March 15, 2017 front page. The regression results from these non-controversial topics were very different from controversial topics. There were very few variables that were significant for either keyword.

Mindfulness keyword. Of the 1000 selected tweets on Muslim on 3-15-2017, 246 had at least one retweet. These tweets were analyzed by BirdIQ, Microsoft Azure Machine Learning Sentiment Analysis, and LIWC. The data collected from this analysis were then examined via linear regression with Retweet count as dependent variable and the 99 variables as independent variables. The regression results had an r square of .548 and was significant at $p < .001$. Most of the variables were not found to be significant, with only 6 variables significant at $p < .10$ as shown in table 1.

In general, retweet popularity was influenced by very different factors than the controversial tweets. In fact the only variables significant at $p < .05$ were question marks, dashes, and numbers. A review of the actual tweets shows that lists and quotes and the main reason for these correlations. In makes sense that a concept of mindfulness would lend itself to quotes of mindfulness gurus and lists of self-help guidelines. The other variables significant at $p < .10$ were risk, adverbs, and tone. Higher risk (also includes doubt), less adverbs, and a more somber tone correlated with higher retweets.

Table 3. Mindfulness data

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-12.664	43.594		-0.291	0.772
Number	11.283	2.428	0.471	4.647	0
QMark	-5.133	1.225	-0.275	-4.189	0
Dash	-6.469	3.044	-0.166	-2.125	0.035
Risk	12.912	7.074	0.216	1.825	0.07
Adverb	-5.793	3.328	-0.225	-1.74	0.084
Tone	-0.563	0.335	-0.285	-1.679	0.095

The final data analyzed were the tweets based on the non-controversial keyword topics, snowstorm, selected from the New York Times March 15, 2017 front page.

Snowstorm keyword. Of the 1000 selected tweets on Muslim on 3-15-2017, 213 had at least one retweet. These tweets were analyzed by BirdIQ, Microsoft Azure Machine Learning Sentiment Analysis, and LIWC. The data collected from this analysis were then examined via linear regression with Retweet count as dependent variable and the 99 variables as independent variables. The regression results had an r square of only .297 and an overall significance of .998 or not significant. Most of the variables were separately not found to be significant, with no variables significant at $p < .05$ and only 6 variables significant at $p < .10$. In an attempt to develop a regression equation with a significance of $p < .05$, the variables from the first attempt (5 variables, authentic, assent, comma, WC, and sad) were test against the dependent variable. When this was attempted, none of the remaining dependent variables were significant at $p <$

.10. This suggests that a non-controversial news event such as a snowstorm does not have any sentiment or linguistic variables that affect the number of retweets.

Overall, it was found that the variables that affected controversial topics were very different from those that may or may not affect non-controversial topics. In fact, the only variable significantly affecting both a controversial (Muslim) topic and a non-controversial topic was Tone. A sad or somber tone for both Muslim tweets and mindfulness tweets significantly increased retweets.

DISCUSSION AND CONCLUSION

Overall, this study has demonstrated a series of important results. First it defines and present a detailed process for capturing Twitter feeds and analyzing these tweets through a series of available software. BirdIQ was used to capture tweets based on a keyword and compute retweets and other relevant information from the Twitter datastream. Though not significant in any of the analysis, Microsoft Azure Machine Learning Sentiment Analysis was used to determine the degree of polarity (positive or negative sentiment) in tweets. Finally, LIWC was used to develop detailed linguistic variables for each of the mined tweets. Finally, all these variables were tested against Retweet count from BirdIQ to see if there were significant independent variables that affected the dependent variable.

Results of the study found significant variables that affected the number of retweets for controversial topics. For Trump retweets, the regression results had an r square of .636 and was significant at $p < .001$. Twenty variables were significant at $p < .10$ and are shown in table 1. In general, retweet popularity was influenced by many tangible and basic needs such as money, family, sex, home, leisure and affiliation. But the two variables that had the largest impact were Dictionary words and Function words.

Muslim keyword regression results had an r square of .567 and was significant at $p < .001$. Twenty-seven variables were significant at $p < .10$ and are shown in table 1. In general, retweet popularity was influenced by somewhat different factors than the Trump tweets. Here we see anger, tone, clout, certainty, cause, negation as significant factors. This suggests a more emotional affect to retweeting texts with keyword Muslim. The highest impact variables were I, differ, prep, cognitive process, and insight. The high use of I in Muslim retweets suggests a much more personal aspect to the religious related retweets. The cognitive processes and insight variables suggest a more thoughtful and reasoning approach. In general, it seems that the Trump retweets were more simplistic and direct but the Muslim retweets were more thoughtful and personal.

There were areas of commonality for the controversial Trump and Muslim tweets. The following variables were significant for both keywords: affiliation, family, function words, I, insight, negate, sexual, and you. Many of these variables suggest personal attitudes that may appeal on an emotional level with individuals. Both were influenced by belonging (family, affiliation) and both were influenced by personal pronouns (I and you). Both also had higher retweets for simpler tweets (function words) but both also were affected by insight and sex. One major difference here though was that Trump retweets had lower insight and Muslim retweets had higher.

Much different results were found for non-controversial topics. In general, neither snowstorm nor mindfulness showed textual variables that influenced retweet count at $p < .05$. The only variable significantly affecting both a controversial (Muslim) topic and a non-controversial topic was Tone. A sad or somber tone for both Muslim tweets and mindfulness tweets significantly increased retweets.

This process was used for only 1000 tweets for each of 4 keywords on one particular day. This is a limitation of the study but the procedures can be used to more extensively test sentiment and linguistic features across a broader sample size. Other limitations include the imprecise nature of sentiment analysis. Though the Microsoft Azure analysis was used, the degree of certainty of results is somewhat questionable due to the short nature of the 140 character tweet. Nevertheless, this research presents a preliminary review of the nature of tweets in the US today and presents an approach that other researchers can expand and confirm via other studies. We believe the paper does present a significant unique research contribution by exploring the sentiment and linguistic analysis for different types of Twitter topics, controversial and non-controversial. Further research is needed to confirm our findings.

REFERENCES

- Back, M. D., Küfner, A. C., & Egloff, B. (2011). Automatic or the people? Anger on September 11, 2001, and lessons learned for the analysis of large digital data sets. *Psychological Science*, 22(6), 837-838.
- Coppersmith, G., Harman, C., & Dredze, M. (2014, May). Measuring Post Traumatic Stress Disorder in Twitter. In ICWSM.
- Eguchi, K. & V. Lavrenko, (2006). *Sentiment retrieval using generative models*, In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 345–354.
- González-Ibáñez, R., Muresan, S., & Wacholder, N. (2011, June). Identifying sarcasm in Twitter: a closer look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2* (pp. 581-586). Association for Computational Linguistics.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- North Concepts (2017). BirdIQ software.
- Pak, A., & Paroubek, P. (2010, May). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *LREC 10*(2010).
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.
- Pennebaker Conglomerates (2015). *LIWC How It Works*. <http://liwc.wpengine.com/how-it-works/>
- Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015). *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates (www.LIWC.net),
- Pennebaker, J., Boyd, R., Jordan, K., and Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX. University of Texas at Austin.
- Peslak, A. R. (2016)a. Do Social Networking Sites Privacy Policies Differ? A Linguistic Analysis of the Ten Most Popular Social Networking Sites. *Issues in Information Systems*, 17(1).
- Peslak, A. R. (2016)b. *Sentiment Analysis and Opinion Mining: Current State of the Art and Review of Google and Yahoo Search Engines' Privacy Policies*. In Proceedings of the Conference for Information Systems Applied Research ISSN (Vol. 2167, p. 1508).
- Robinson, R. L., Navea, R., & Ickes, W. (2013). Predicting final course performance from students' written self-introductions: A LIWC analysis. *Journal of Language and Social Psychology*, 0261927X13476869.
- Sexton, J. B., & Helmreich, R. L. (2000). Analyzing cockpit communications: the links between language, performance, error, and workload. *Journal of Human Performance in Extreme Environments*, 5(1), 6.
- Smith, K. (2016). *44 Twitter Statistics for 2016*. *Brandwatch*. <https://www.brandwatch.com/blog/44-twitter-stats-2016/>
- Stieglitz, S., & Dang-Xuan, L. (2012, January). Political communication and influence through microblogging--An empirical analysis of sentiment in Twitter messages and retweet behavior. In System Science (HICSS), 2012 45th Hawaii International Conference on (pp. 3500-3509). IEEE.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM, 10*(1), 178-185.

APPENDIX A

Dependent and Independent variable table

Variable	Meaning or examples
RetweetCount	Number of retweets of a specific tweet
UserFollowersCount	Number of users following the person who initially posted the tweet
UserFollowingCount	Number of other people the user who made the tweet is following
UserTweets	Numner of tweets user has made
UserFavoritesCount	Number of favorites the user has designated
SentN	Overall sentiment of tweet (0 negative to 1 positive)
WC	Word count of tweet
Analytic	Tweet reflects logical thinking versus narrative
Clout	Confident (high score) versus Tentative (low score)
Authentic	Honest versus Guarded
Tone	Upbeat versus sad
WPS	Words per sentence
Sixltr	Number of six letter words
Dic	Number of words in LIWC dictionary (suggests ease of read)
Function	it, to, no, very
Pronoun	I, them, itself
Ppron	I, them, her
I	I, me, mine
We	we, us, our
You	you, your, thou
Shehe	she, her, him
They	they, their, they'd
Ipron	it, it's, those
Article	a, an, the
Prep	to, with, above
Auxverb	am, will, have
Adverb	very, really
Conj	and, but, whereas
Negate	no, not, never
Verb	eat, come, carry
Adj	free, happy, long
Compare	greater, best, after
Interrog	how, when, what
Number	second, thousand
Quant	few, many, much
Affect	happy, cried
Posemo	love, nice, sweet
Negemo	hurt, ugly, nasty
Anx	worried, fearful
Anger	hate, kill, annoyed
Sad	crying, grief, sad
Social	mate, talk, they
Family	daughter, dad, aunt
Friend	buddy, neighbor
Female	girl, her, mom
Male	boy, his, dad
Cogproc	cause, know, ought
Insight	think, know
Cause	because, effect

Issues in Information Systems

Volume 18, Issue 2, pp. 48-58, 2017

Discrep	should, would
Tentat	maybe, perhaps
Certain	always, never
Differ	hasn't, but, else
Percept	look, heard, feeling
See	view, saw, seen
Hear	listen, hearing
Feel	feels, touch
Bio	eat, blood, pain
Body	cheek, hands, spit
Health	clinic, flu, pill
Sexual	horny, love, incest
Ingest	dish, eat, pizza
Drives	
Affiliation	ally, friend, social
Achieve	win, success, better
Power	superior, bully
Reward	take, prize, benefit
Risk	danger, doubt
Focuspast	ago, did, talked
Focuspresent	today, is, now
Focusfuture	may, will, soon
Relativ	area, bend, exit
Motion	arrive, car, go
Space	down, in, thin
Time	end, until, season
Work	job, majors, xerox
Leisure	cook, chat, movie
Home	kitchen, landlord
Money	audit, cash, owe
Relig	altar, church
Death	bury, coffin, kill
Swear	fuck, damn, shit
Netspeak	btw, lol, thx
Assent	agree, OK, yes
Nonflu	er, hm, umm
Filler	I mean, you know
AllPunc	
Period	
Comma	
Colon	
SemiC	
QMark	
Exclam	
Dash	
Quote	
Apostro	
Parenth	
OtherPunc	