Twitter Analytics-Based Assessment: Are the United States Coastal Regions Prepared for Climate Change?

Abstract—**According to the U.S. National Climate Assessment, the Southeast Coast and Gulf Coast of the United States are particularly susceptible to sea level rise, heat waves, hurricanes and lower accessibility to clean water due to climate change. Preparation for climate change consequences can only occur with conversation, which is a method of bringing awareness to the issue. Over the past decade, social media has taken over the spectrum of information exchange in the United States. Social Network Analysis (SNA) is the practice of analyzing trends in volume and opinion of a population of social media users. Twitter, one popular social media platform, is one of the largest microblogging sites in the world, and it provides an abundance of data related to the trending topics such as climate change. In this work, Twitter analytics is performed on the data generated from Twitter users in the United States, who were talking about climate change, global warming and/or CO2, from July 2016 to June 2017. Specifically, a comparative sentiment analysis on the coastal U.S. regions was conducted to recognize which region(s) is/are falling behind on the conversation about climate change and to understand the trends in opinion about climate change over time. The results determined that the southeast coast of the United States is deficient in their discussion about climate change compared to the other coastal regions. Igniting the conversation about this issue in these regions will mitigate the disasters due to climate change by increasing awareness in the people of these regions so they can properly prepare.**

Keywords—climate change; social media analysis; sentiment analysis; twitter data; US coastal regions

I. INTRODUCTION

In the past decade, the percentage of American adults who use social media has increased by over 800% [14]. Also in the past decade was born a social media giant, Twitter.com – a massive portal of information where over 500 million tweets are posted every day [15]. A tweet is a 140-character post that the Twitter user uses to express any opinion, idea or fact that comes to mind. Due to the volume of conversation that occurs via Twitter and other social media platforms, the field of social network analysis (SNA) has surfaced as a useful tool in many disciplines [6]. SNA takes the data that is generated by the dialog that occurs through social media to gain knowledge about current trends, public opinion, demographic information, etc.

We would like to thank the *National Science Foundation* (REU Award AGS 1560210 and MRI Award AGS 1624068) for funding support. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

According to the U.S. National Climate Assessment, the coastal regions of the United States are particularly susceptible to sea level rise, heat waves, hurricanes and less accessibility to clean water due to climate change. This is because of the extreme variation of topography in these two regions [12]. Preparation for climate change consequences can only occur with conversation, which is a method of bringing awareness to the issue. Social media is the medium that today's society utilizes to communicate with each other [4].

The Pew Research Center for Excellence in Journalism in 2009 revealed that social media posts related to climate change are much more prominent recently. It has been confirmed that while many posts on Twitter are simply conversational, it does also serve as a news network [16] and a venue for users to communicate political issues [7]. One method that can be used to analyze the information that is exchanged through social networking, is sentiment analysis through text mining. Sentiment analysis is used to understand levels or trends in the emotion behind a message over a period of time. Text mining is the process of deriving insights from a body of text [6].

While climate change is predicted to affect the earth as a whole, one of the main consequences of global warming/climate change is sea level rise, coastal regions are particularly susceptible to the consequences of climate change [11]. In this work, SNA was performed using data from Twitter to examine the trends in sentiment and volume of tweets related to climate change in the coastal regions of the United States. This analysis was done through statistical analysis and text mining techniques. A regional comparative analysis was executed to visualize the current status of the Twitter conversation on climate change of each of the coastal regions in the United States.

Several techniques such as the usage of Naïve Bayes [17], usage of tree kernel for efficient feature engineering [18], utilization of linguistic features and existing lexical resources [19] have been applied to perform sentiment analysis on the twtter data. Usage of twitter data to perform sentiment analysis and opinion mining has been carried out for several applications and scenarios such as breast cancer treatment experiences and perceptions [20], popularity of city locations [21], US presidential elections [22], stock market predictions [23], adverse drug reaction analysis [24], retail store analysis [25], and examination of Alzheimer's disease stigma [26].

A similar study that was done on tweets about climate from 2008-2014 suggested that the majority of climate change tweets were from activists, rather than deniers, which indicates that Twitter is an effective mechanism for increasing awareness of the issue [4]. To our knowledge, there is no work in the literature [2,3,5,8,9, 10 and 13] that deals with the study of conversation on the effects of climate change in the coastal regions of the United States.

To address the existing limitations in the literature, the contributions of this study are to: (1) visualize the level of awareness about climate change in each coastal region of the United States based on the volume and sentiment of the tweets generated in that region, (2) understand the trends in sentiment of the tweets related to climate change over a one-year period. (3) to identify the region(s) lacking in the conversation to boost recognition and preparation for the consequences of climate change.

The rest of the paper is organized as follows. Section 2 describes the data and methodology used for the experimentation and analysis. Section 3 describes the volume comparison and sentiment analysis experimentation results. Section 4 discusses the sentiment analysis results. Section 5 describes the future wok recommendations and finally section 6 provides the summary and conclusion.

II. Data and Methodology

In this study, 238,570 geotagged tweets, sourced within the continental U.S., were collected using the Twitter Stream Application Programming Interface (API), comprising a twelve-month period from July 1 2016, until June 30th, 2017. These tweets were extracted using spatiotemporal criteria and keywords (climate change, global warming and/or CO_2) in the tweet message and hashtags. The locational accuracy of a geotagged tweet depends on how a Twitter user shares his/her location when posting a tweet. The location can be shared in the format of place names (e.g., country, state, and city) or the exact latitude and longitude (point-level, determined by the device's GPS or other signals such as cell tower).

A. Data

Preliminary social network analysis was performed on the tweets that originated in the United States, were posted between July 2016 and June 2017, contained one or more of the keywords: "climate change", "global warming", "CO2", and included the user's geographical location, which is also referred to as "geotagged". The tweets that met these criteria were stored in a data set containing 38 variables that include the user's location, the message they posted (tweet) and the date/time stamp.

B. Methodology

Observations in the data were sorted in order to perform regional comparative analysis of the coastal regions of the United States, the Southeast Coast, Northeast Coast, West Coast, and Gulf Coast (as shown in **Figure 1**), to compare tweet volume and sentiment in each region.

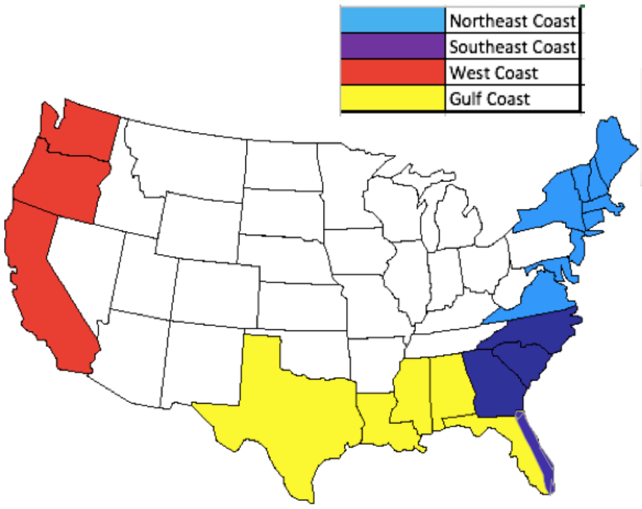

Figure 1. Regions of Comparative Analysis

2.2.1 Volume Comparison

Initially, the raw volume of tweets generated from each region was compared. This simply analyzed the number of tweets in the data set that originated in each region. Then, to ensure unbiased data, the number of tweets from a particular region was normalized by the corresponding population. This normalization provides a representation of the amount of discussion per capita. For each region, the normalized volume of tweets was analyzed to identify trends and related (by percentage) with respect to the national dialogue on the topic of climate change.

2.2.2 Sentiment Analysis

To analyze the feeling behind the text, which is the tweet, a sentiment analysis was performed. While different methodologies exist [6], the analysis in this study was grounded on the "sentiment score" of the processed tweets based on the bing.com dictionary of the most commonly used words in the English language to express feeling (see Listing 1). Each word in this dictionary has an associated binary sentiment value. If the word expresses positive emotion, it has a value of 1. If the word expresses negative emotion, it has a value of -1. The algorithm that was implemented matched the words in the conversation with the dictionary database and assigned a score. Every tweet then received a total score, which was a sum of the values that were associated with the matched words.

For example, a tweet of "if climate change is a crisis why aren't we seeing more sacrifice from people who think it's a problem" received a score of -2. The words "crisis" and "problem" each have a score of -1, which yields the score of -2. The rest of the words in that tweet were classified as neutral (not strongly affiliated with a positive or negative attitude), so this tweet receives a score of -2. As shown in Listing 1, the algorithm then summed up all of the tweets' scores for each region. The overall sum for each region calculated a net score that revealed if the majority of the population was using positive or negative words to express their opinion of the issue of climate change.

```
library("tm")
library("syuzhet") #necessary packages for this method or
sentimentscore<-function(regtweets,mon,txtcol) #regtweets
                                                #mon was
                                               #txtcol is
{
  regmon<-regtweets[c(which(regtweets$month == mon)),]
  regmontidy<-regmon$txtcol
  #convert all text to lower case
  regmontidy<- tolower(regmontidy)
  # Replace blank space ("rt")
  regmontidy<- gsub("rt", "", regmontidy)
  # Replace @UserName
  regmontidy<- gsub("@\\w+", "",regmontidy)
  # Remove punctuation
  regmontidy<- gsub("[[:punct:]]", "",regmontidy)
  # Remove links
  regmontidy<- gsub("http\\w+", "",regmontidy)
  # Remove tabs
  regmontidy<- gsub("[ |\t]{2,}", "", regmontidy)
  # Remove blank spaces at the beginning
  regmontidy<- gsub("^ ", "", regmontidy)
  # Remove blank spaces at the end
  regmontidy<- gsub(" $", "",regmontidy)
  score<-sum(get_sentiment(regmontidy,"bing"))
  return(score)
}
```

Listing 1- Algorithm for Sentiment Analysis

2.3 Experiment Design

The tools used in this Twitter analytics work were Matlab and RStudio. Matlab, a mathematical modeling software was used to do most of the plotting and visualizations. RStudio, a statistical analysis software, was used to do the volume and sentiment analysis. It has been demonstrated earlier in the literature that 'R' language and RStudio can be applied to perform sentiment analysis [1]. RStudio facilitated the text mining portion of this study using the "tm" package. This is also the package that contained the dictionary of words that were utilized during the sentiment analysis. As part of the experimentation, 238,570 geotagged tweets, sourced within the continental US were loaded into R for data preparation and analysis.

III. Results

A. Volume Analysis Results

Results showed that the Southeast Coast of the U.S. only contributed to 9% of the overall conversation happening about climate change in the United States Coastal regions while the other 3 regions each make up around 21 to 40 percent of this conversation.

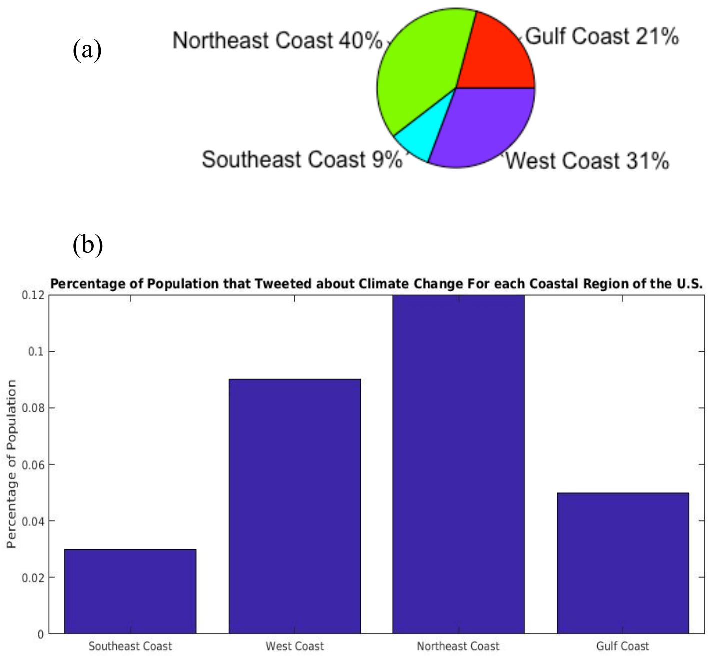

Figure 2. Contribution of Each Region to the Coastal U.S. Conversation on Climate Change. (a) The pie chart shows the volume of Tweets between 07/16 and 06/17 about climate change as partitioned by U. S. coastal regions. (b) Comparison of the Percentage of the population of each region.

In order to create an even field of data, the results were normalized by population size of each region. The number of tweets from each region was divided by the population in each region to give the percentage of the population in that region that tweeted about climate change between July 2016 and June 2017. This visualization continues to reveal a shortage in overall discussion about global warming in the southeast coast.

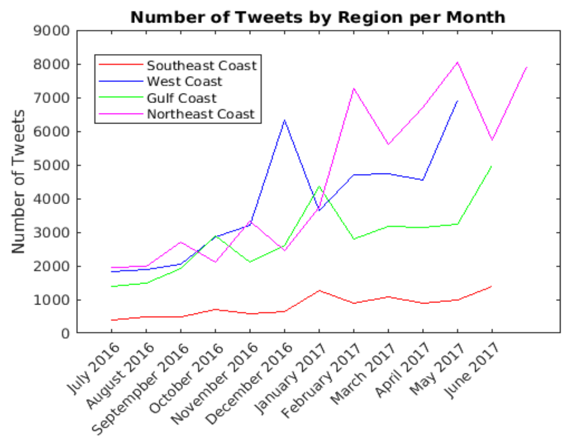

Figure 3. Comparison of the Volume of Tweets Generated by each Region per Month

The volume of tweets from each region that were posted each month was plotted over the one-year period. Figure 3 shows that while the Gulf Coast, West Coast and Northeast Coasts are tweeting at similar volumes each month, the Southeast

Coast was lacking in the volume of tweets about climate change that are being generated each month.

Once again, to ensure that the dearth of conversation in the southeast coast, which is what is revealed by the visualization, is not due to higher populations in the other regions compared to the population of the southeast coast, each month was normalized by the population in each region. Figure 4 displays that the southeast coast is still lagging behind the other regions. However, it is apparent that the percentage of the population that is tweeting is on the rise for every region.

Although each region is generating an increasing amount of tweets each month, not all of the regions are keeping up with the national accumulation of Twitter conversation about climate change. Figure 5 reveals that the southeast and gulf coast have a dwindling contribution to the overall climate change tweet volume in the U.S. At the rate that these two regions are declining, the southeast coast's involvement in the conversation will be practically nonexistent in the next couple of decades. The Gulf Coast, whose participation is deteriorating 10 times faster than the southeast coast, will theoretically make no significant contribution to the Twitter conversation on climate change by the year 2027.

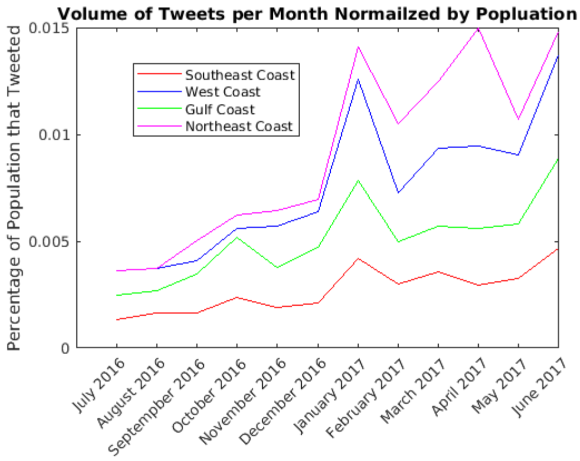

Figure 4. Comparison of the % of Population that Tweeted about Climate Change each Month

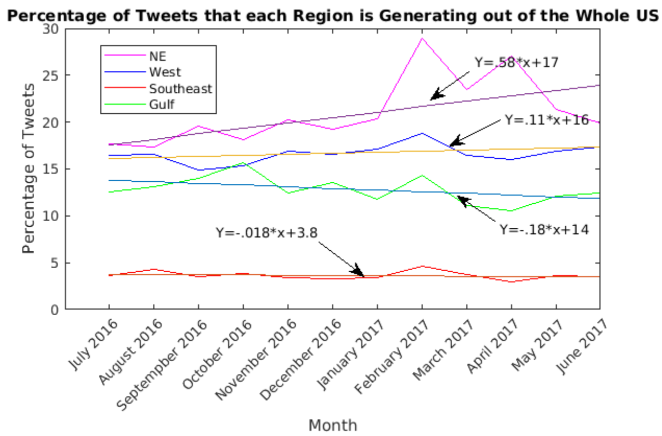

Figure 5. Comparison of the Percentage of the National Tweets about Climate Change that are Generated from each Region.

B. Sentiment Analysis Results

The overall sentiment score is representative of the emotion that was expressed by the majority of people in that region. This bar graph below shows the positive score, which is the sum of all the positive sentiment value for the tweets in each region, the negative sentiment score, and the net score. Figure 6 shows the composition of each net score.

The sentiment scores for each month were then plotted (Figure 7) to see how the net sentiment score varied over a period of time. An obvious downward spike in the level of sentiment occurred in January of 2017, which is hypothetically caused by a large amount of people angry about the more political facet of climate change based on the review of a few tweets from that month. The sharp upwards spike in April for the Northeast Coast is currently unexplained, but could be due to a warmer weather that was occurring during this month. Figure 7 also revealed that the southeast coast (red) has a very small magnitude year-round relative to the other regions that show obvious spikes during certain months. The fact that the southeast coast's sentiment score revolves around zero for the entire year suggests that the tweets that were generated from that region were evenly split between positive and negative sentiment. The smaller population size of the southeast coast is partly to blame for the smaller degree of sentiment score, but Figure 6 reveals that the southeast coast's net score is much lower in magnitude than the other regions, due to the almost identical composition between positive and negative sentiment.

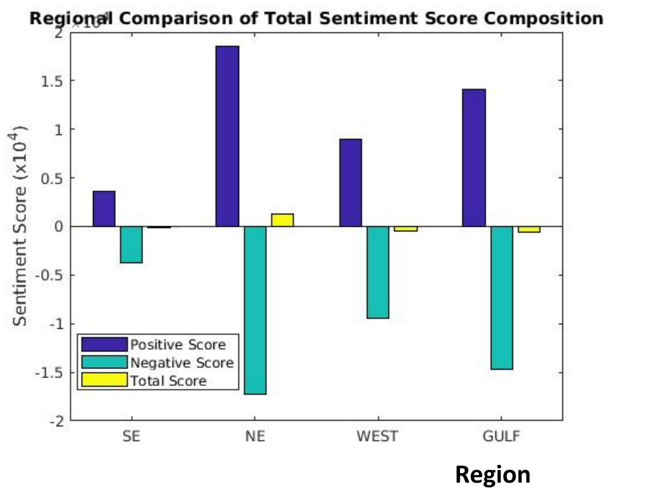

Figure 6. Composition of the Net Sentiment Score that each Region Received over a Period of one year.

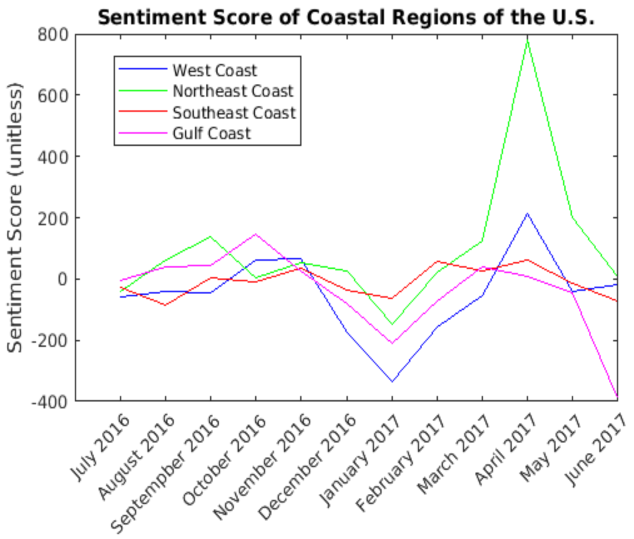

Figure 7. Composition of how the sentiment score for each region changes every month

IV. Discussion On Sentiment Analysis

The sentiment analysis in this study is accurate when observing trend in extreme disgust towards climate change, as it was obvious in January of 2017 in Figure 7**.** However, as shown in Figure 6, the majority of the scores that are calculated are comprised of very positive and negative scores that are close in magnitude. The limitation of this analysis lies in the dictionary that was used to assign values to each word. Had the dictionary's values been specific to what a certain word would mean in the context of the conversation of climate change, the results would have been different. For instance, a tweet that stated, "Climate change is very real, everyone should believe in it" would be given a positive score due to the words "real" and "believe". Alternatively, a tweet that said, "Government funding for climate change research is horrible, climate change is a serious problem" would have a negative score due to the words "horrible" and "serious problem". However, both of those tweets indicate a person who is concerned about climate change. These tweets should be compared to the tweets that say things like "climate change is a hoax" or "who cares about global warming?" which indicate someone who is not concerned about climate change.

Conversely, the sentiment analysis technique that was implemented in this study was very accurate when identifying trends in the negative sentiment, especially when the tweets are directed more towards environmental policy and less towards the climate itself. In that case, the words used to express emotion about the policy are obviously negative or positive. The work done on the sentiment analysis has plenty of room for improvement, but the results indicate that the tweets of the southeast coast are divided between positive and negative emotions.

V. Future Work

While this work primarily analyzed the tweets pertaining to the coastal states of the Unites states, in the future, we plan to extend this analysis to global scale, primarily comparing the volume of tweets among the top English speaking countries such as United States, Canada, England and Australia. With respect to the actual text mining methods, we plan to expand to the text mining methods such as topic modeling, in addition to the sentiment analysis.

At a more local level, this study could be improved and continued by taking into account regional cultures and likelihood that people in each region will tweet or handle issues, such as climate change, in another manner. The age and socioeconomic status of the tweeters in each region would give valuable information. In addition, looking at the number of unique twitter users in each region would account for frequent tweeters to give a more accurate volume of the population tweeting in each region.

VI. Conclusions

Climate change has been forecasted to have global impacts, but due to sea level rise, and increasing storm surges, coastal regions are especially at risk [11]**.** This study aimed to quantify the amount of conversation that was taking place about climate change to identify if there was a specific coastal region that was lacking in awareness of the issue compared to the others. The comparative analysis on volume of tweets from each region led to the conclusion that the Southeast Coast region of the United States is having less conversation than other coastal regions of the United States. The only way for a region to be prepared for the possible disasters that climate change has the potential to impose is to be aware of the issue. Awareness comes from discussion through the media. Social media is a powerful medium to inform and converse, and it can be an effective tool to increase the awareness about issues in society, such as climate change.

To increase the awareness of climate change in a specific region, the conversation has to be sparked initially, which is the issue addressed in the volume analysis – more conversation is imperative to increasing awareness of this issue. Nevertheless, the conversation that is started must be conducive to the process of making people understand the seriousness of the problem. Conducive conversation occurs when the majority of a population is in agreement about a subject, rather than constantly having half of a population express one opinion, while the other expresses the opposite. In the other coastal regions of the United States, there are certain months where there are spikes upwards or downwards in sentiment score. The spikes indicate that the majority of that population in that region is in agreement about the subject, at least for that period of time. This is what the southeast coast needs more of a louder, more influential conversation that will raise awareness and help the region prepare for the impact that climate change will trigger.

ACKNOWLEDGMENT

We would like to acknowledge the help of Vivian Turner, Varavut Limpasuvan, and Todd Rhodes for their editorial help and stimulating discussions.

REFERENCES

[1] A. Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. "Sentiment analysis of twitter data." In Proceedings of the workshop on languages in social media, pp. 30-38. Association for Computational Linguistics, 2011.

[2] Arun, K., A. Srinagesh, and M. Ramesh. "Twitter Sentiment Analysis on Demonetization tweets in India Using R language." International Journal of Computer Engineering in Research Trends 4, no. 6 (2017): 252-258.

[3] An, X., A. R. Ganguly, Y. Fang, S. B. Scyphers, A. M. Hunter, and J. G. Dy (2014). Tracking climate change opinions from twitter data. In Workshop on Data Science for Social Good.

[4] Auer, M. R., Y. Zhang, and P. Lee (2014). The potential of microblogs for the study of public perceptions of climate change. Wiley Interdisciplinary Reviews: Climate Change, 5(3), 291-296.

[5] Boykoff, M.T. Who speaks for the climate?: Making sense of media reporting on climate change. Cambridge University Press; 2011

[6] E. Clark, Eric M., Ted James, Chris A. Jones, Amulya Alapati, Promise Ukandu, Christopher M. Danforth, and Peter Sheridan Dodds. "A Sentiment Analysis of Breast Cancer Treatment Experiences and Healthcare Perceptions Across Twitter." arXiv preprint arXiv:1805.09959 (2018).

[7] Cody, E. M., A. J. Reagan, L. Mitchell, P. S. Dodds, and C. M. Danforth (2015). Climate change sentiment on twitter: an unsolicited public opinion poll. PloS one, 10(8), e0136092.

[8] Collomb, A., C. Costea, D. Joyeaux, O. Hasan, and L. Brunie, (2015). A Study on Sentiment Analysis: Methods and Tools. International Journal of Science and Research (IJSR),4(12), 287-292. doi:10.21275/v4i12.nov151832

[9] Jungherr, A., H. Schoen, and P. Jürgens (2015). The Mediation of Politics through Twitter: An Analysis of Messages posted during the Campaign for the German Federal Election 2013. Journal of Computer-Mediated Communication, 21(1), 50-68. doi:10.1111/jcc4.12143

[10] Kirilenko, A. P. and S. O. Stepchenkova (2014). Public microblogging on climate change: One year of Twitter worldwide. Global Environmental Change, 26, 171-182.

[11] Kirilenko, A. P., T. Molodtsova, and S. O. Stepchenkova (2015). People as sensors: Mass media and local temperature influence climate change discussion on Twitter. Global Environmental Change, 30, 92-100.

[12] Kouloumpis, Efthymios, Theresa Wilson, and Johanna D. Moore. "Twitter sentiment analysis: The good the bad and the omg!." Icwsm 11, no. 538-541 (2011): 164.

[13] Mandel, B., A. Culotta, J. Boulahanis, D. Stark, B. Lewis, and J. Rodrigue (2012, June). A demographic analysis of online sentiment during hurricane irene. In Proceedings of the Second Workshop on Language in Social Media (pp. 27-36). Association for Computational Linguistics.

[14] Mcgranahan, G., D. Balk, and B. Anderson, (2007). The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones. Environment and Urbanization,19(1), 17-37. doi:10.1177/0956247807076960

[15] Melillo, J. M., T. C. Richmond, and G. W. Yohe, Eds., 2014: Climate ChangeImpacts in the United States: The Third National Climate Assessment. U.S. Global Change Research Program, 841 pp. doi:10.7930/J0Z31WJ2.

[16] M. Moh, Teng-Sheng Moh, Yang Peng, and Liang Wu. "On adverse drug event extractions using twitter sentiment analysis." Network Modeling Analysis in Health Informatics and Bioinformatics 6, no. 1 (2017): 18.

[17] Olteanu, A., C. Castillo, N. Diakopoulos, and K. Aberer (2015). Comparing events coverage in online news and social media: The case of climate change. In Proceedings of the Ninth International AAAI Conference on Web and Social Media (No. EPFL-CONF-211214).

[18] N. Oscar, Pamela A. Fox, Racheal Croucher, Riana Wernick, Jessica Keune, and Karen Hooker. "Machine learning, sentiment analysis, and tweets: an examination of Alzheimer's disease stigma on Twitter." Journals of Gerontology Series B: Psychological Sciences and Social Sciences 72, no. 5 (2017): 742-751.

[19] A. Pak, and P. Paroubek. "Twitter as a corpus for sentiment analysis and opinion mining." In LREc, vol. 10, no. 2010. 2010.

[20] Pew Research Center (2017, January 12) Social Media Fact Sheet Retrieved July 22nd, 2017 from http://www.pewinternet.org/fact-sheet/social-media/

[21] Rao, Tushar, and Saket Srivastava. "Analyzing stock market movements using twitter sentiment analysis." In Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), pp. 119-123. IEEE Computer Society, 2012

[22] Smith, K. (2016, November 16) Astonishing Twitter Stats for 2016, Retrieved July 22nd, 2017 from https://www.brandwatch.com/blog/44-twitter-stats-2016/

[21] Stoimenov, Leonid, and Aleksandra Đukić. "Using Sentiment Analysis of Twitter Data for Determining Popularity of City Locations." ICT Innovations 2016 (2018): 156.

[22] Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle." In Proceedings of the ACL 2012 System Demonstrations, pp. 115-120. Association for Computational Linguistics, 2012.

[23] Yates, D. and S. Paquette (2011). Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. International Journal of Information Management, 31(1)

[25] Yujiao, Li, and Hasan Fleyeh. "Twitter Sentiment Analysis of New IKEA Stores Using Machine Learning." In International Conference on Computer and Applications. 2018.