Sentiment Analysis of User-Generated Content from Twitter, Facebook, and YouTube on the Topic of Mental Health: Research Proposal

Emily Tyger

SODA 496

December 14, 2020

Introduction

Brands are learning how to precisely manage their content on social media in order to keep up with the growing technologies in computer and data science that allow for the detailed analysis of online content relating to their brand. When a brand successfully communicates their product or image through an advertisement or customer interaction on social media, the consumer is more likely to make a purchase, and it improves the brand's relationship with the user/consumer (Agnihotri et al., 2016). A brand can learn how to effectively communicate their message on social media by studying their data through a sentiment analysis to better understand the opinions and emotions of their customers who have a social media presence (Seabrook et at., 2018). I propose research on the topic of mental health with a dataset comprised of user-generated content, known to include many emotions and opinions, in order to observe patterns or differences in the sentiment. By analyzing users' sentiment on extreme topics such as mental health, across different platforms, I have the opportunity to expand upon previous research that is lacking in cross-platform analysis consisting of more than 2 platforms.

Research Questions and Purpose Statement

The purpose of the explanatory research proposed in this report is to determine how sentiment regarding the topic of mental health varies across social media platforms Twitter, Facebook, and YouTube, concerning user-generated content, such as posts and comments. This research aims to answer three main research questions:

1. How does sentiment on the topic of mental health, from user-generated content, vary across Twitter, Facebook and YouTube?
2. How are likes, shares, and other attributes of user-generated content on the topic of mental health associated to the sentiment score; is there variation across Twitter, Facebook and YouTube?
3. What is the distribution of educational, entertainment, and neutral sentiment scores across content generated by users of different ages; is there variation across Twitter, Facebook and YouTube?

Conceptualization

User-generated content is defined as content that is, “not created within professional routines or practices” which “can be seen as the sum of all ways in which individuals make use of social media” (Kaplan & Haenlein, 2010). It is relevant to study user-generated content, the way social media users interact on the platform, because it is a way to analyze the user’s emotions and opinions on sensitive topics. Social media, defined by Tuten and Solomon’s (2017) textbook, is stated to be, “the online means of communication, conveyance, collaboration, and cultivation among interconnected networks of people, communities, and organizations enhanced by technological capabilities”. The social media platforms this report is concerned with includes Twitter, Facebook, and YouTube (Tuten & Solomon, 2017). It is important to note that user-generated content does not include advertisements nor does it include posts or comments by organizations, businesses, or institutions.

Twitter is a social media platform used as an outlet for news, information, opinions, complaints, and details about daily life, featuring posts with a maximum of 200-characters that can be shared, liked, or sent via message. Posts can also include hyperlinks, photo and video media, surveys, tags to other users’ profiles on the site, and location tags as well (Boyd et al., 2010). Facebook is defined as a social network site which differs from Twitter as Facebook users are more likely to use the platform to maintain connections, communicate with friends, and share personal details (Debatin et al., 2009). The greatest difference between YouTube and the other social media platforms considered in this report, is the content is comprised primarily of videos. Users can post professional or user-generated videos, and interact through likes, dislikes, shares, comments, and subscribing to other users. Interactions on the platform are not unlike others, despite YouTube’s classification as a content community (Burgess & Green, 2018). Due to these differences, analyzing the way user’s experience and user-generated content sentiments vary across these social media platforms has been used in various contexts including previous research on mental health, immunization, and audio-visual sentiment analyses.

The research proposed will utilize a sentiment analysis to answer the research questions. A sentiment analysis is defined as the process of analyzing a segment of text to automatically detect if it contains emotional or opinionated information, and then determining the polarity of the user-generated content (Paltoglou & Thelwall, 2012). The goal is to examine whether there is sentiment occurring in social media content that is associated with a user's pattern of information sharing on the platform. In some cases, sentiment is easily detected due to the report that emotionally charged tweets are more likely to be retweeted, and the retweet occurs much faster than it would for neutral tweets (Stieglitz & Dang-Xaun, 2013). A sentiment analysis is a strong approach because it can be applied to subjectivity detection of content sentiment as well as polarity classification of content sentiment; the proposed research will be applied to polarity classification in user-generated content about mental health topics.

User-generated content is increasing at an exponential rate, full of crucial information for researchers to analyze and Paltoglou and Thelwall (2012) highlight the importance of improving methods of conducting a sentiment analysis due to the lack of a "golden standard" in the domain of informal communication. Movie, or general product reviews, are more standard and easier to analyze than informal communication, and as informal data is exponentially increasing, this poses a challenge for researchers (Paltoglou & Thelwall, 2012). A sentiment analysis may be utilized to classify content as educational or entertainment for the purpose of this research proposal. Educational content may include words that point to resources, information, or support to other users. The goal of content created for entertainment is to poke fun at or make light of this difficult topic, or be relatable to other users. Previous research has touched on cross-platform analysis as well as the sentiment of content, which was used to develop the theory behind this proposal.

Literature Review

Research from 2017 worked with big data to analyze 176 million tweets on topics related to depression or suicide in order to identify deviations from predicted trends in communication about mental health (McClellan et al., 2017). In order to conduct the analysis, researchers utilized an autoregressive integrated moving average model with which they observed heightened interest in mental health related topics following national or international events; these events include World Suicide Prevention Day and comedian Robin William's suicide. The research calls attention to studies that have analyzed Twitter content on similar topics such as flu season, drug and alcohol addictions, physical activity, vaccinations, cancer and obesity but regardless, there is a lack of research detailing the mental health of social media users through user-generated content. A stronger understanding of consumer's emotions online allows brands to better market to their audience to drive sales and strengthen the brand-customer relationship (Dhar & Chang, 2009).

User-generated sentiment analysis has also been used to research mental health on YouTube by understanding the sentiment behind pro-anorexia and anti-pro-anorexia content. One study aims to analyze emotional reactions to pro-anorexia content on YouTube with a collection of 395 videos and 12,161 comments. The sampling method started with the 50 most popular YouTube channels on the topic of anorexia and the final dataset included over 6 million total views across the videos collection, and 8000 comments by unique users (Oksanen et al., 2015). A sentiment analysis was conducted, followed by an ordinary least squares regression to test the strength of the sentiment on positive and negative observations along with video likes and dislikes for videos that are pro- and anti-pro-anorexia. The results of the analysis are statistically significant enough to conclude that user-generated content with positive sentiment scores are correlated to the videos with the highest interactions as far as views, likes, comments, and shares (Oksanen et al., 2015). This means that the videos supporting positive mental health information is getting more attention, which combats opposition group's negative presence.

A similar research study conducted by Yiannakoulias, Slavik, and Chase (2019) analyzes how users express pro- and anti-vaccine sentiment on YouTube. This research is followed by a discussion of a similar pro- anti-immunization sentiment analysis on Facebook by Hoffman, Felter, Kar-Hai, Shensa, and Hermann (2019). It is beneficial for healthcare officials to understand the nature of how information about health is being shared on social media platforms Facebook and YouTube in order to better understand how to control the spread of rumors and misinformation. From 220 videos collected by analyzing Google Trends for "flu" and "vaccine"-related topics, researchers were able to describe immunization pro- and anti- expressions, to then compare the expressions with that of pro- and anti-immunization sentiments (Yiannakoulias et al., 2019). The goals of a similar Facebook sentiment analysis are to characterize the users who post anti-immunization behavior on Facebook, the information these users convey, and the spread of this user-generated content. To create the dataset, researchers selected 197 Facebook users who responded to pro-immunization content with an anti-immunization user-generated response (Hoffman et al., 2019). Since the amount of Facebook (Hoffman, 2019) users is close to the size of the YouTube (Yiannakoulias, 2019) dataset, the results of the 2 studies indicate that YouTube is a popular platform for viewing user-generated content on immunization, however, Facebook is a stronger platform for creating and sharing user-generated content on immunization. These analyses are beneficial to institutions and companies looking to successfully leverage their data.

Sentiment analysis has been evolving on social media to include research on audio-visual sentiments as well. Research conducted by Wollmer, Weninger, Knaup, Schuller, Sun, and Sagae (2013) looks at movie reviews in the form of YouTube videos in order to expand upon the lack of research concerning expanding the application of sentiment analysis to modalities such as speech, gesture, and facial expressions. The topic of movie reviews is strong for this research because sentiment analysis originated with movie and general product reviews, as they contain less common language that is challenging to interpret. Researchers use this to their advantage as they build a dictionary of audio-visual reviews that are matched with original movie review dictionaries of the text-based sentiments. The results provide an impression of the corresponding accuracies for classification of a dataset of 370 review videos as visual and linguistic information cannot be automatically obtained without manual transcription of online reviews at the time of the analysis (Wollmer et al., 2013).

More has been researched on the individual social media platforms, as well as the comparison of Facebook and Twitter. Previous literature (2019) observes user-generated content on Facebook such as likes, comments, shares, and reactions, and how this impacts the intensity of the sentiment from a Facebook diabetes page. To conduct the sentiment analysis, machine learning algorithms including random forest, support vector machine, and logistic regression were used to improve the accuracy of results that may indicate the user’s behavior to have a positive impact on the intensity of the sentiment (Kaur et al., 2019). The results show that different combinations of three dimensions did not make a substantial impact on the outcome, and the distribution of positive, negative, and neutral sentiments is relatively equal. The combination of likes, comments, and shares, is slightly greater than the other combinations, indicating this trio has the strongest impact on sentiment intensity. The literature that cross-examines social media applications tends to assess Facebook and Twitter together more frequently than other platforms. Marshall, Ferenczi, Lefringhausen, Hill, and Deng (2020) present an analysis of how Facebook-only users differ from Twitter users. The research presented calls attention to differences in attributes such as intellect, creativity, career promotion, frequency, social connection, and personal achievement. Twitter users result in higher scores for intellect and creativity than Facebook-only users. The majority of Twitter users like to read about intellectual pursuits and posted more frequently; it is to be noted that Twitter is designed to accommodate more frequent posts than Facebook. Facebook-only users scored significantly higher than Twitter users with content about personal achievement and social connection (Marshall et al., 2020).

The work by Oksanen et al., (2015) analyzing pro- and anti-pro-anorexia YouTube content utilizes a smaller dataset than the mental health Twitter analysis (2017). The results of these findings, in which the most positive user-generated content occurs in response to the anti-pro-anorexia videos, are similar to the solution to the proposed research question of "How are likes, shares, and other attributes of user-generated content on the topic of mental health associated to the sentiment score; is there variation across Twitter, Facebook and YouTube?" I will expand upon this research by not only analyzing the YouTube sentiment of content with a large dataset, I will additionally compare the results to the results of a similar analysis on Twitter and Facebook as well. It is interesting to note the following among the research analyzing the anti-immunization compared to the anti-pro-anorexia YouTube user-generated content: anti-pro-anorexia, healthier content, was more popular on YouTube and this content received a more positive sentiment score however, the anti-immunization, less healthy, content was more popular on YouTube and Facebook. This may be attributed to content including negative sentiment and dramatized emotions receiving higher views and shares as a whole across all social media platforms.

Research by Wollmer et al., addresses how sentiment analysis research over the last decade has primarily been studied to classify reviews of products and services, along with more recently, user-generated content on social media. The researchers highlight the lack of attention to sentiment analysis regarding audio-visual content because media of this type is growing at an exponential rate as technology further advances to handle larger, more advanced data. While I will not be utilizing audio-visual sentiment analysis in my research, I note that this is a shortcoming of my analysis as YouTube content is primarily audio-visual, and without this analysis I will be limited to the user-generated content solely from commenters, rather than directly from the source, the creator.

Previous research has yet to adequately address YouTube in comparison to other social media platforms including Twitter and Facebook. While much research has addressed the comparison of Twitter and Facebook, YouTube is a growing platform with a unique subset of social media users as the literature has suggested and it is beneficial for brands to understand this relationship as well. Due to previous research by Marshall, Ferenczi, Lefringhausen, Hill, and Deng that highlights major differences between Twitter and Facebook users, including YouTube users in my analysis will further strengthen the relevance of my findings which can be used to greater the understanding of variance in sentiment in user-generated content.

Theory

In order to gain a more complete understanding of how users interact on social media, it is crucial to not only understand what topics or words are trending, but the way in which people are talking about the topic. I propose an analysis of content on the same topic as it appears on different social media sites in order to understand how different ways of discussing the topic of mental health impacts the performance and popularity of the content.

Operationalization

In order to analyze multiple platforms, operationalization specific to each platform must be considered individually. First, user-generated content appears on Facebook in the form of a post whereas on YouTube user-generated content appears as videos including a description. For the purpose of this research, user-generated content on YouTube will be collected in the form of video descriptions due to the lack of technology supporting audio-visual analysis. On Twitter, user-generated content comes in the form of a character limited post however, original tweets and replies to other posts appear on user's timelines in an almost identical form which must be considered at the time the content is sampled and collected.

Similar to the form of user-generated content, likes and shares differ in terminology across all three platforms as well. On YouTube, individuals have the opportunity to like or dislike a video and both counts are to be included in the final dataset, as well as the ratio of positive to negative reactions. This is similar to how a Facebook user "reacts" to a post, which complicates the collection of this independent variable because reactions include love, like, laugh, wow, sad or angry. For the purpose of this research proposal, a "like" on Facebook will include likes, loves, laughs and wow reactions. The negative reactions will be included in the dataset similar to YouTube in order to calculate the positive to negative ratio. On Twitter, a like comes in the form of a favorite and is not as complex as the other platforms. Fortunately, shares across all platforms appear in a nearly identical fashion, although shares on Twitter are considered "retweets." One important note about shares on Facebook and YouTube compared to Twitter, is that content may be shared to other followers on the platform, *or* outside of the platform in the case a post is sent via SMS or other outlet. Retweet counts do not include shares outside of the platform which may skew results when comparing shares across social media networks.

One of the most complex differences across Facebook, YouTube, and Twitter for the purpose of this research is the attribute for creator followers. Followers on Facebook are known as "friends" which requires a reciprocal relationship, meaning that a friend request is sent by one user to another and in order to become a follower, the other must first accept the request. In turn, the one who received and accepted the request will also be following the user who sent the request; this is the only social media platform that holds a one to one relationship like this. Twitter on the other hand, allows users to follow any account without that account having to return the follow which may be considered a one to zero *or* one relationship. YouTube is the most complex of all as the majority of YouTube users do not have to be signed into an account to view a video, but they must be signed in to subscribe to another user. This being the case, users who subscribe to a content creator typically do not have subscribers themselves, creating a many to one relationship. This being the case, it could be beneficial to include a feature for the user-generated content's views in order to compare this number to the creator's number of subscribers.

The last attribute to be included in the dataset will hold the age of the content creator when available. While each platform requires a date of birth upon the creation of an account, the user can choose to make this information private or public. When this information is made available, the date of birth will be scraped from the user profile and converted into the current age of the user to be used in the analysis. The user's age is a critical part of this analysis however, due to the potential lack of availability, it may be difficult to gain a complete representation of the population.

Hypotheses

1. I expect to find a stronger relationship between the age of the user posting, rather than the likes on the post, and the label of the post.
2. I anticipate the most prevalent label on each social media platform will average more likes per user-generated content.
3. I expect as the age of the user increases, their content is more likely to be classified as educational.

Population and Sampling

The population of interest for this research includes Twitter, Facebook, or YouTube users who created content regarding the term "mental health" who are active from 2015-present. The population excludes professional accounts from businesses or organizations because posts from these accounts are not considered user-generated content as defined for the purpose of this research. Sampling will be conducted three times, once across each platform, collecting 50,000 posts on each social media site for a total dataset of 150,000 observations. For each social media network, sampling is conducted by querying for user-generated content that includes terms related to mental health including but not limited to: mental health, anxiety, depression, and therapy. In order to ensure the dataset is comprised of user-generated content alone, posts from verified accounts will be excluded from sampling as the majority of these accounts are from professionals, organizations, or celebrities which are unnecessary for the purpose of this research. Additionally, sampling for each platform will not consider if a user has an account on any of the other platforms; the purpose of this explanatory research is to understand how users interact on each platform individually. Only accounting for users that have an account on multiple platforms could skew the results and not show an accurate representation of each platform.

Data Sources

Data will be collected from Facebook, YouTube, and Twitter directly through each platform's API. By utilizing Python packages for each of these platforms including Python-Facebook, Python-YouTube, and Python-Twitter, the data will be collected uniformly and the results of the collection will be returned in similar formats, making the data easy to work with. From previous experience with Python-Twitter, I am familiar with the json dictionaries that are returned and I will use this experience to my advantage for Python-Facebook and -YouTube as well. A disadvantage to this approach is the lack of information made available through the YouTube API, as well as the difficulty in acquiring a developer account through Google in order to access the API. In the case that I am not able to collect the necessary YouTube data, I can use previous experience from scraping web pages through their html format in order to collect the description, likes/dislikes, views, and information on the creator as well.

Proposed Analytical Method

When working with text data, preprocessing is a critical step before any analysis can be conducted due to the complications involved with text analysis. Posts on Facebook, YouTube, and Twitter can include characters and information that does not enhance the analysis such as emojis or hyperlinks which must be removed through preprocessing. Another difficulty when working with text as data includes the high dimensionality of the individual words used in the post. The first way to reduce the dimensionality of text data is to remove stop words that are present in all posts and do not enhance the analysis such as articles "a", "an" and "the" and including other terms such as "it", "and", "be", etc. I can utilize the Python package stop-words to find a dictionary of words to exclude. In order to clean the data so that it is uniform and only includes the necessary information, other preprocessing steps include converting the post to lowercase so that "Anxiety" and "anxiety" are not considered separate terms as the computer understands them to be. Additionally, I can utilize stemming and lemmatization to generate the root of the words used in the post so that, following the same example, "anxiety" and "anxious" are reduced to their common base and therefore, will be considered the same word for the purpose of this research.

Final preprocessing steps will include the conversion of date of birth into the age of the content creator utilizing the package datetime, which makes this feature much easier to work with and analyze. I will also need to create additional columns in the dataset for ratios such as the positive to negative reactions on Facebook as well as the subscriber to views ratio on YouTube. Before the data frames for each platform are combined into one final dataset, it is necessary to include a column to identify the source (platform) of the post. Last, I will need to determine the best way to handle null values as they appear in the dataset after sampling.

The first step in this analysis is to label the observations as “educational” or “entertainment” by nature of the words included in the post after preprocessing. As the dataset contains no labels, the sentiment analysis will require unsupervised learning because this method includes no prior assumptions about the data and will allow the model to learn the structure of the data itself. The main idea behind this approach is that educational and entertainment posts will include different words (outside of the sampled terms regarding mental health), but posts identified as educational will include words more similar to each other than words included in entertainment posts. A clustering algorithm will be best for this sentiment analysis in order to divide the posts.

Once the dataset is labelled, I will utilize an ordinary least squares regression to draw samples from the population to estimate other features that may be present as hypothesized. The goal of a regression analysis is to minimize discrepancy between estimated and predicted values that result from the sentiment analysis through the other features in the dataset. The regression results will identify the strength of the relationship between the label of the post, the dependent variable, and the independent variables: likes, shares, followers, and age. The regression coefficient on the independent variables will indicate the degree to which a one unit increase in this feature impacts the label of the post, as well as a measure of how statistically significant this result is in terms of what I have hypothesized.

Conclusion

As companies strive to keep up with growing technologies in computer and data science, it is crucial to not only understand what topics or words are trending, but the way in which people are talking about the topic in order to get the most out of your data. Brands are learning how to precisely manage their content on social media in order to improve their relationship with consumers by effectively communicating their message. I propose an analysis of content on the topic of mental health across three major social media platforms in order to better understand how users express emotions and opinions on a sensitive topic that can be used to learn how users interact with each platform differently. By analyzing users' sentiment on extreme topics such as mental health, I have the opportunity to expand upon previous research that is lacking in cross-platform analysis consisting of more than 2 platforms.

By design, Facebook, YouTube, and Twitter are completely different platforms, therefore it makes sense that their users may greatly vary as well. Twitter is used as an outlet for information and news, featuring a character limit that highlights the app's fast pace while Facebook is known for making connections and sharing more personal information. YouTube is comprised of less creators than the other platforms by nature of the user-generated content in the form of videos, and is known by users as a content community. If a company wants to advertise their product, based on the design of the platforms alone, it would not make sense to create one advertisement to use for all three social media networks.

Utilizing a large dataset of 150,000 observations I aim to add to previous research that lacks cross-platform analysis through a sentiment analysis followed by an ordinary least squares regression. By classifying posts as educational or entertainment, I will be able to gain a better understanding of the way users discuss the same topic across platforms, and determine which sentiment is more popular on each platform, including the number of posts for each classification on each platform, the number of likes and shares each classification receives, and the age of the creator to determine if there is a generational divide in users. The research proposed in this report is designed to be easily reproduced for future analysis and improvements. This research can be expanded by including other platforms such as Instagram or TikTok that are comprised of media and audio-visual data that is more challenging to analyze with current technology.

Citations

Agnihotri, R., Dingus, R., Hu, M. Y., Krush, A.T. (2016). Social media: Influencing customer satisfaction in B2B sales. *Industrial Marketing Management*, *53*, pp. 172-180. https://doi.org/10.1016/j.indmarman.2015.09.003

Boyd, D., Golder, S., & Lotan, G. (2010, February). *Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter*. System Sciences: 43rd Hawaii International Conference on System Sciences, Hawaii.https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1109%2FHICSS.2010.412

Burgess, J., & Green, J. (2018). *YouTube: Digital Media and Society Series*. Polity. https://books.google.com/books?hl=en&lr=&id=mg1rDwAAQBAJ&oi=fnd&pg=PT5&dq=Jean+Burgess,+Joshua+Green+YouTube+(Digital+Media+and+Society+Series)+Polity+Press,+Cambridge,+UK+(2009)&ots=RCnIMDm8vM&sig=9E15JJLPLCii4CCs2tty3q3vuHI#v=onepage&q=Jean%20Burgess%2C%20Joshua%20Green%20YouTube%20(Digital%20Media%20and%20Society%20Series)%20Polity%20Press%2C%20Cambridge%2C%20UK%20(2009)&f=false

Debatin, B., Lovejoy, J. P., Horn, A., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication, 15*(1), pp.83-108. https://doi.org/10.1111/j.1083-6101.2009.01494.x

Dhar, V., & Chang, E. A. (2009). Does change matter? The impact of user-generated content on music sales. *Journal of Interactive Marketing*, *23*(4), pp. 300-307. https://doi-org.ezaccess.libraries.psu.edu/10.1016/j.intmar.2009.07.004

Hoffman, B. L., Felter, E. M., Kar-Hai, C., Shensa, A., Hermann, C., et al. (2019). It's not all about autism: The emerging landscape of anti-vaccination sentiment on Facebook. *Vaccine, 37*(16), pp. 2216-2223. https://doi.org/10.1016/j.vaccine.2019.03.003

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, *53*(1), pp. 59-68. https://doi-org.ezaccess.libraries.psu.edu/10.1016/j.bushor.2009.09.003

Kaur, V., Balakrishnan, V., Rana, O., & Sinniah, A. (2019). Liking, sharing, commenting, and reacting on Facebook: User behaviors' impact on sentiment intensity. *Telematics and Informatics*, *39*, pp. 25-36. https://doi-org.ezaccess.libraries.psu.edu/10.1016/j.tele.2018.12.005

Marchall, T. C., Ferenczi, N., Lefringhausen, K., Hill, S., & Deng, J. (2020). Intellectual, narcissistic, or Machiavellian? How Twitter users differ from Facebook-only users, why they use Twitter, and what they tweet about. *Psychology of Popular Media*, *9*(1), pp. 14-30. https://doi.org/10.1037/ppm0000209

McClellan, C., Ali, M. M., Mutter, R., Kroutil, L., & Landwehr, J. (2017). Using social media to monitor mental health discussions—evidence from Twitter. *Journal of the American Medical*

Informatics Association, *24*(3), pp. 496-502. https://doi-org.ezaccess.libraries.psu.edu/10.1093/jamia/ocw133

Oksanen, A., Garcia, D., Sirola, A., Nasi, M., Kaakinen, M., Keipi, T., & Rasanen, P. (2015). Pro-Anorexia and anti-pro-anorexia videos on YouTube: Sentiment analysis of user responses. *Journal of Medical Internet Research*, *17*(11). https://doi.org/10.2196/jmir.5007

Paltoglou, G., & Thelwall, M. (2012). Twitter, MySpace, Digg: Unsupervised sentiment analysis in social media. *ACM Transactions on Intelligent Systems and Technology*. https://doi-org.ezaccess.libraries.psu.edu/10.1145/2337542.2337551

Seabrook, E. M., Kern, M. L., Fulcher, B. D., & Rickard, N. S. (2018). Predicting depression from language-based emotion dynamics: Longitudinal analysis of Facebook and Twitter status updates. *Journal of Medical Internet Research*, *20*(5). https://www.jmir.org/2018/5/e168/

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media-sentiment of microblogs and sharing behavior. *Journal of Management Information*, *29*(4), pp. 217-248. https://doi-org.ezaccess.libraries.psu.edu/10.2753/MIS0742-1222290908

Tuten, T. L., & Solomon, M. R. (2017). *Social Media Marketing*. Sage Journal. https://books.google.com/books?hl=en&lr=&id=XQg_DwAAQBAJ&oi=fnd&pg=PT15&dq=social+media+&ots=tQ8ZxV2kqO&sig=dm4N2GfO4lDdsx81xGgIFB6jgWQ#v=onepage&q&f=false

Wollmer, M., Weninger, F., Knaup, T., Schuller, B., Sun, C., Sagae, K., et al. (2013). YouTube move reviews: Sentiment analysis in an audio-visual context. *IEEE Intelligent Systems*, *28*(3), pp. 46-53. https://doi-org.ezaccess.libraries.psu.edu/10.1109/MIS.2013.34

Yiannakoulias, N., Slavik, C., Chase, M. (2019). Expressions of pro- and anti-vaccine sentiment on YouTube. *Vaccine*, *(37*)15, pp.2057-2064. https://doi-org.ezaccess.libraries.psu.edu/10.1016/j.vaccine.2019.03.001