NLP Model for Effect of Social Media on Health

Brianni Duncan
bduncan6@gmu.edu
Research Contribution: Brianni Duncan has contribution in both research and coding part of this work.

***Abstract*—Social Media is a tool that offers users entertainment, creative expression, networking opportunities, more accessible access to an abundance of resources. However, there is an alarming increase in mental illnesses associated with the high use of social media. Given that mental health is an imperative aspect of an overall healthy lifestyle, it is vital to consider how millions of people utilize and rely on social media applications daily, making them vulnerable to mental illnesses associated with social media use. In addition to this, a particular demographic may be at high risk: a target age group, people of color, underprivileged communities, and a particular gender.**

***Index Terms*—Social Harm, Social Benefit, Social media, NLP**

I. INTRODUCTION

One of the most prevalent age groups that utilize social media is teenagers. With social media existing as a platform for competition for popularity, teenagers respond to societal beauty norms, popular, attractive, and appealing. A study conducted by Maria Clark in 40+ Frightening Social Media and Mental Health Statistics shows a linkage where the suicide rates of teenagers have risen from 2011 by 150 percent, and the depression rate has grown by 112 percent. It is pretty noticeable that the rate is increasing per the introduction of social media networking platforms such as Facebook, WhatsApp, Instagram. In an effort to respond, Instagram has suppressed likes to curb the comparisons and hurt feelings associated with attaching popularity to sharing content. Nevertheless, this made the most negligible impact on teenagers' mental health changes, and rates have not improved effectively. [1] [2] [3] [4] [5].

Social media engagement can make people feel closer to society and their near and dear, but this also can lead to FOMO (Fear of Missing Out) when left alone, potentially causing disruptions in mental health. Ironically, social media activates the brain's reward center by releasing dopamine, which is recognized as a happy chemical. Dopamine releases during pleasurable experiences, such as eating a favorite food, interacting with loved ones, and exercising; when it is lacking, this leads to Anxiety and Depression.

Similarly, the social media activities are designed in such a way that it is addictive and brings Anxiety and Depression when it is unavailable. However, where there can be positive effects of social media, there are even more negative impacts on children and adults. [6]

sentiment and emotion analysis helps Mental health detection on social media [7]–[10] [1]. One of the most prevalent age groups that utilize social media is teenagers. With social media existing as a platform for competition for popularity, teenagers respond to societal beauty norms, popular, attractive, and appealing. A study conducted by Maria Clark in 40+ Frightening Social Media and Mental Health Statistics shows a linkage where the suicide rates of teenagers have risen from 2011 by 150 percent, and the depression rate has grown by 112 percent. It is pretty noticeable that the rate is increasing per the introduction of social media networking platforms such as Facebook, WhatsApp, Instagram. In an effort to respond, Instagram has suppressed likes to curb the comparisons and hurt feelings associated with attaching popularity to sharing content. Nevertheless, this made the most negligible impact on teenagers' mental health changes, and rates have not improved effectively. [1]

II. LITERATURE REVIEW

We want to consider previous studies conducted by researchers over the past 5-10 years to assist in navigating our research. Furthermore, we intend to observe how analysts, medical professionals, mental health advocates, and social networking representatives respond to the phenomena. Thus, literature research is essential to our study.

Numerous articles exist suggesting a correspondence between social media use and mental health risks. Social Bots can affect mental health of the society [11]–[15]. This literature review will evaluate two articles: Social Media Use and Its Connection to Mental Health: A Systematic Review and Social Media and Adolescents' and Young Adults' Mental Health. We assess these particular articles for several reasons. The first being to adopt data collection and processing methods from knowledgeable researchers. The second reason is to assist in defining a research problem and proposing a solution. Lastly, our group intends to seek a clear understanding of how best technology utilizes in research strategies.

Transfer Learning for NLP models in Social media helps to detect these trends [16] [17]–[34]. [16]Our research group adopted a similar method in our systematic study by extracting one hundred scholarly articles highlighting the relationship between social media usage and mental health. The two central databases accessed were Google Scholar and George Mason University's online library. Collectively, we gathered twenty-five articles per student, focusing on content containing similar keywords of "social media," "social networking," "mental health," "mental disorder," and "technology and mental health." The second article we are reviewing, Social Media and Adolescents' and Young Adults' Mental Health, studies the relationship between social media use among adolescents and how their mental health is in decline. This article supports the first composition conclusions claiming that 25% of adolescent users believe their social media habits negatively impact their mental health. A strong link also exists between sleep patterns and high social media usage. Poor sleep patterns commonly lead to depression and anxiety, two mental illnesses prominently cited in the first article. [16]

A. Problem Investigation

While social media has brought countless social benefits, the increased interconnectivity has introduced unforeseen social issues. Among younger people in particular, where social media use is ubiquitous, there is concern that social media produces an adverse effect on mental health. According to an article regarding Social Media published by the National Center for Health Research, ”With 13% of 12–17-year-olds reporting depression and 32% reporting anxiety, mental illness is a concern for adolescent health. It is a concern for young 2adults as well since 25% of 18–25-year-olds report having some form of mental illness”. [35] [36] [37] [38] [39] [40] [41]

III. Approach

Analytics begins with raw data, so our main priority was finding reliable data sources to support our thesis. We have gathered articles from various resources and used R Studio for web scraping and converting into CSV files. Once all our data converts into a single structured format, we have performed data pre-processing on the extracted data. Data is pre-processed and cleaned using Python, eliminating any anomalies, including removing the stop words, connective words, and other special characters or missing data. Since the data gathered comes from various resources, we will be creating a standard format to analyze data further and gain valuable insight. [42]

IV. Data Pre-Processing

We have gathered the articles that incorporate the terms ”Social Media” and ”Mental Health” in the title for pre-processing the data. These articles were gathered from various sources such as GMU Resources (Online Database and libraries). Once we gathered all the related articles we used R Studio for web scraping using the HTML Method. We used multiple libraries such as Rvest and PDFtools in R studio inorder to extract data from websites as well as pdf.

Codes were repeated for all the websites and PDFs that we chose based on the category that we selected. The above code stores the text from the website and pdf in the list format in R environment, so for our analysis purposes we had to convert the list into a dataframe. Once all the lists were converted into dataframes, data frames were exported to our local hard drive for the purpose of combining all the data frames. Using MS Excel CSV format for the exported files to combine data. Once all the data frames were combined, cleaning processes started where any empty row was taken out, any row which had irrelevant data was deleted such as references, author names,

Fig. 1. Figure III

title etc. That took most of our time as we went in manually to verify that once data is loaded it does not have any unnecessary text that can affect the end result which is referred to as noise. Even after converting our dataframes into csv after cleaning we realised that the code that we used stored each line from the text was separated by comma not each word so we used MS excel feature data tab, text to column feature is used. through which each word was separated into its own column. Since there were still a lot of unnecessary words in our data such as stop words we used MS excel feature find and replace on all the stop words that were present through which a lot of unnecessary data was deleted we finally had our clean data ready for Text Analysis.

V. Text Analysis

Once we had created a clean dataset, we were ready to begin analysis of the text. Importing the dataset into Monkeylearn, a text analysis tool that uses machine learning, we were able to generate meta-data about the combined dataset. First was a simple word count to show which areas were focused by the researchers (Figure 1). From the word count, it is simple to see which areas were commonly addressed in the papers: social media, mental health, trangender people, and more. Many of the studies were focused on adolescents and other young people, also demonstrated in the word count. A word count such as this one allowed us to quickly see which areas of classification to focus on moving forward.

VI. Conclusion

Variety of symptoms and issues, such as minority or transgender vulnerabilities, and concerns about children and young teens than we originally anticipated. We also found cyberbullying and addiction to social media to have a wide footprint across the scholarly literature on the subject. These findings are key to our classification and further work we are doing through machine learning.

References

[1] E. Mir, C. Novas, and M. Seymour, "Social media and adolescents' and young adults' mental health." https://www.center4research.org/social-media-affects-mental-health/, 2021.

[2] World Wide Web Foundation, "History of the web." https://webfoundation.org/about/vision/history-of-the-web/, 11 April 2021.

[3] Harvard Medical School Affiliate, "The social dilemma: Social media and your mental health." https://www.mcleanhospital.org/essential/it-or-not-social-medias-affecting-your-mental-health, February 9, 2021.

[4] U.S.Gov, "U.S. and World Population Clock." https://www.census.gov/popclock/, 21 April 2021.

[5] Dr. C. George Boeree, "Abraham maslow." http://webspace.ship.edu/cgboer/maslow.html, 2006.

[6] M. Clark, "40+ frightening social media and mental health statistics." https://etactics.com/blog/social-media-and-mental-health-statistics, 12 November 2020.

[7] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "Emotion detection of textual data: An interdisciplinary survey," in *IEEE 2021 World AI IoT Congress,AIIoT2021*, 2021.

[8] B. Dao, T. Nguyen, S. Venkatesh, and D. Phung, "Nonparametric discovery of online mental health-related communities," in *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1–10, 2015.

[9] B. Dao, T. Nguyen, S. Venkatesh, and D. Phung, "Effect of social capital on emotion, language style and latent topics in online depression community," in *2016 IEEE RIVF International Conference on Computing Communication Technologies, Research, Innovation, and Vision for the Future (RIVF)*, pp. 61–66, 2016.

[10] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "A survey on concept-level sentiment analysis techniques of textual data," in *IEEE 2021 World AI IoT Congress,AIIoT2021*, 2021.

[11] A. H. Yazdavar, M. S. Mahdavinejad, G. Bajaj, K. Thirunarayan, J. Pathak, and A. Sheth, "Mental health analysis via social media data," in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 459–460, 2018.

[12] Y. Zhao, Y. Guo, X. He, J. Huo, Y. Wu, X. Yang, and J. Bian, "Assessing mental health signals among sexual and gender minorities using twitter data," in *2018 IEEE International Conference on Healthcare Informatics Workshop (ICHI-W)*, pp. 51–52, 2018.

[13] M. Heidari, J. H. J. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter," in *IEEE 2020 International Conference on Data Mining Workshops (ICDMW), ICDMW 2020*, 2020.

[14] M. Heidari and J. H. Jones, "Using bert to extract topic-independent sentiment features for social media bot detection," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, pp. 0542–0547, 2020.

[15] M. Heidari, J. H. J. Jones, and O. Uzuner, "An empirical study of machine learning algorithms for social media bot detection," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[16] C. Berryman, C. J. Ferguson, and C. Negy, "Social media use and mental health among young adults," *Psychiatric Quarterly*, vol. 89, pp. 307–314, Nov. 2017.

[17] M. M. Aldarwish and H. F. Ahmad, "Predicting depression levels using social media posts," in *2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS)*, pp. 277–280, 2017.

[18] K. Katchapakirin, K. Wongpatikaseree, P. Yomaboot, and Y. Kaewpitakkun, "Facebook social media for depression detection in the thai community," in *2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 1–6, 2018.

[19] M. Heidari and S. Rafatirad, "Using transfer learning approach to implement convolutional neural network model to recommend airline tickets by using online reviews," in *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA*, pp. 1–6, 2020.

[20] H. Yang, C. He, H. Zhu, and W. Song, "Prediction of slant path rain attenuation based on artificial neural network," in *2000 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 1, pp. 152–155 vol.1, 2000.

[21] A. Zahura and K. A. Mamun, "Intelligent system for predicting suicidal behaviour from social media and health data," in *2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*, pp. 319–324, 2020.

[22] L. Liu, B. Li, I.-M. Chen, T. J. Goh, and M. Sung, "Interactive robots as social partner for communication care," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2231–2236, 2014.

[23] M. Heidari and S. Rafatirad, "Bidirectional transformer based on online text-based information to implement convolutional neural network model for secure business investment," in *IEEE 2020 International Symposium on Technology and Society (ISTAS20), ISTAS20 2020*, 2020.

[24] L. Dobrescu, S. Obreja, M.-C. Vochin, D. Dobrescu, and S. Halichidis, "New approaches for quantifying internet activity," in *2019 E-Health and Bioengineering Conference (EHB)*, pp. 1–4, 2019.

[25] A. J. Majumder, J. W. Dedmondt, S. Jones, and A. A. Asif, "A smart cyber-human system to support mental well-being through social engagement," in *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 1050–1058, 2020.

[26] J. Parraga-Alava, R. A. Caicedo, J. M. Gómez, and M. Inostroza-Ponta, "An unsupervised learning approach for automatically to categorize potential suicide messages in social media," in *2019 38th International Conference of the Chilean Computer Science Society (SCCC)*, pp. 1–8, 2019.

[27] M. Heidari, J. H. J. Jones, and O. Uzuner, "Misinformation detection model to prevent spread of the covid-19 virus during the pandemic," 2021.

[28] F. Murtagh, "Analysing activities, contextualized for general health, depression and demographics," in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 547–554, 2018.

[29] C.-H. Tai, Z.-H. Tan, and Y.-S. Chang, "Systematical approach for detecting the intention and intensity of feelings on social network," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 987–995, 2016.

[30] M. Heidari, S. Zad, B. Berlin, and S. Rafatirad, "Ontology creation model based on attention mechanism for a specific business domain," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[31] J. Du, Y. Zhang, C. Tao, and H. Xu, "A pilot study of mining association between psychiatric stressors and symptoms in tweets," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1254–1257, 2017.

[32] M. Heidari and S. Rafatirad, "Semantic convolutional neural network model for safe business investment by using bert," in *IEEE 2020 Seventh International Conference on Social Networks Analysis, Management and Security, SNAMS 2020*, 2020.

[33] C. A. V. Palattao, G. A. Solano, C. A. Tee, and M. L. Tee, "Determining factors contributing to the psychological impact of the covid-19 pandemic using machine learning," in *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 219–224, 2021.

[34] M. Heidari, S. Zad, and S. Rafatirad, "Ensemble of supervised and unsupervised learning models to predict a profitable business decision," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[35] L. Robinson and M. Smith, "Social media and mental health." https://www.helpguide.org/articles/mental-health/social-media-and-mental-health.htm, September 2020.

[36] H. Patel, “An introduction to web scraping using R.” https://www.freecodecamp.org/news/an-introduction-to-web-scraping-using-r-40284110c848/, 24 October 2018.
[37] tutorialspoint, “Python - text processing.” https://www.tutorialspoint.com/python_text_processing/index.htm, 2021.
[38] K. Flerlage, “A starter kit for text analysis in tableau.” https://www.flerlagetwins.com/2019/09/text-analysis.html, 28 September 2019.
[39] R. Thorstad and P. Wolff, “Predicting future mental illness from social media: A big-data approach,” *Behavior Research Methods*, vol. 51, pp. 1586–1600, Apr. 2019.
[40] A. MacCaw, G. Cabane, R. Fishkin, and S. Blum, “[14] text analysis with monkeylearn.” https://monkeylearn.com/.
[41] K. Yap, “Text analysis using excel.” https://www.keithyap.com.au/text-analysis-using-excel/#topic-modelling, 25 November 2016.
[42] S. Jain, “Ultimate guide to deal with text data (using python) – for data scientists and engineers.” https://www.analyticsvidhya.com/blog/2018/02/the-different-methods-deal-text-data-predictive-python/, 27 February 2018.