

Social Media Engagement Of Young People: A Sentiment Analysis Using Natural Language Processing

Elaine Marie Aranda¹, Roy Francis Navea²

¹De La Salle University 2401 Taft Ave., Manila, 0922 Philippines.

²De La Salle University 2401 Taft Ave., Manila, 0922 Philippines.

Abstract

Mental health is an integral part of human existence and is most likely underestimated when it comes to human health. The thoughts that occupy the mind affects how a person feel and it corresponds to action which may have beneficial or harmful results. Thoughts are commonly expressed in words through written or verbal communication. In this study, the sentiments of young people are determined through the responses that they provide as they share their social media engagement experiences. With Natural Language Processing (NLP) and the implementation of Valence Aware Dictionary and s Entiment Reasoner (VADER), word clouds were produced containing significant words that indicate positive and negative sentiments. Results show common words in the positive sentiment as well as in the negative sentiment that support and represent the reality of the experience of young people as they engage in different social media activities.

Keywords Mental health; internet; online experience

Introduction

The impact of the COVID-19 pandemic in 2020 placed a large percentage of the population of the world under strict quarantine measures and forced to stay, work, and study from home. The Internet played a major role in connecting people, both local and abroad, to keep on-track about what is going on in the society specially in business, politics, governance, education, leisure, and entertainment. It is well-known that the Internet is a space of connection between people and virtual contents where one can find them useful or not. Useful contents are followed by subscribers while the useless contents are ignored and yet still impact the viewers one way or another. Contents are viewed in a variety of platforms and social media is one of the popular ones. Both adults and the young people are engaged in one or more social media platforms wherein they check on it every single day. Social media engagement is expressed in terms of the activities done in the platform like posting, sharing of contents, reacting such as liking,

sharing, and commenting, or building a promotional page for information dissemination to build connections, influence, and for some, earn a living.

Literature Review

According to Statista Research Department, the highest audience share of social media in the Philippines belongs to the age bucket between 18 to 24, followed by ages 25 to 34 (Statista Research Department, 2021). The younger age bracket shows that they are either in their college levels or they are starting their professional careers. Social media provides beneficial or adverse impacts to its users. Though it is a neutral platform, it has the capacity to make or break a person, an institution, and any digital entity built upon that platform. On the positive perspective, social media contributes to the technical innovations in industries that includes health, services, tourism, businesses, and social digital communities. It is characterized by the sharing of knowledge and expertise, networking, social branding and influence, and the development of new skills in digital mass communications. Hence, young people and adults are joining these media sites and are interacting with friends, relatives, and new acquaintances. Social media influence human behavior which is even amplified by technological advancements in social interaction, sharing of information, and social feedbacking. This has significantly improved the ways on how instructions are delivered in the educational system, how health services and information are shared in a larger percentage of patients, and how businesses scaled up their potentials in marketing, sales, and customer services. Hence, this changed the way how people behave and interact as compared to a decade before this digital innovation. Nonetheless, the impact of social media adversely affected human health in the physical and mental domains. This includes anxiety, depression, stress, and other health problems which are found to be impacting daily human living (Abbas et al., 2019). Majority of the contents in social media are words or text-based materials in combination with some images or moving pictures. Words, in writing or speech, are one of the expressions of thoughts that relate to human feelings and emotions which leads to actions that could either be beneficial or harmful. If words could only be analyzed and early predictions of actuations can be performed, this could hinder and prevent any negative actuation that a person might commit. One of the measures of words expressed is the sentiment. In this study, sentiment is framed as positive or negative. The sentiments of social media users are analyzed through a sentiment analysis algorithm known as the Valence Aware Dictionary and Entiment Reasoner (VADER) implemented in MATLAB (Hutto & Gilbert, 2015; Moßburger et al., 2020). This is a text- or word- based program that uses a sentiment lexicon with words annotated with sentiment scores between -1 to 1. The objective is to determine the concepts behind the engagement in social media of young people through the texts or words coming directly from their responses through Natural Language Processing (NLP) methods (Patton et al., 2020). NLP is a multidisciplinary field of expertise that merges computer science and computational linguistics which aims to bridge human communication and computer understanding. It is a branch of artificial intelligence that helps computers understand, interpret, and manipulate human language using different algorithms and word processing techniques.

Methodology

The block diagram shown in Figure 1 depicts the sentiment extraction procedures. It is divided into three phases which are composed of data gathering, processing, and interpretation.

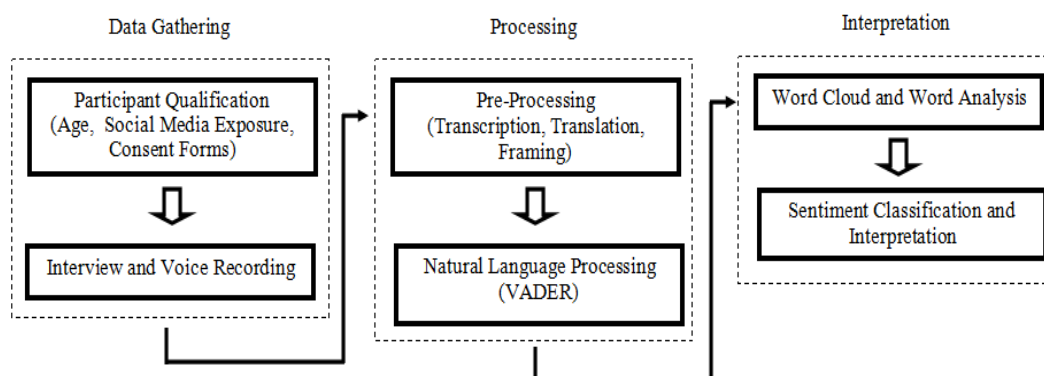


Figure 1: Sentiment extraction procedure

Data Gathering Phase

The key data are the responses of the participants who are in the age bracket of 16 to 24. They should have a minimum social media exposure of 6 hours in a day regardless of whether male or female. Interview protocols include getting of data privacy consents, orientation of the project's background and objectives, and gathering participant's demographics. Their social media usage in the previous week (before the interview) was asked to illustrate their exposure. The interview was done online, and the conversation recordings were kept for transcription. The research questions asked include (RQ1) What is social media for you? (RQ2) Why do you use social media? and, (RQ3) What are the social media contents that make you engaged? A free flow of words was recorded and gathered for sentiment analysis.

Processing Phase

The recorded conversations were transcribed word per word and were translated to the English language. Manual transcription and translation were performed to preserve the context of the conversations (Tho et al., 2021). The responses were framed and indexed in preparation for NLP using the VADER algorithm. The VADER algorithm is used to evaluate sentiment in tokenized text by running the vader Sentiment Scores function in MATLAB. It returns sentiment scores for tokenized documents by aggregating individual token scores, adjusted according to the algorithm rules, and then normalized between -1 and 1. The function can discard single tokened characters that are not present in the sentiment lexicon (MathWorks, 2021). The VADER algorithm also utilizes word lists like Boosters, Dampeners, and Negations that modify the scores of proceeding words in the text. This algorithm is limited to the English language only thus, responses are translated accordingly.

Interpretation Phase

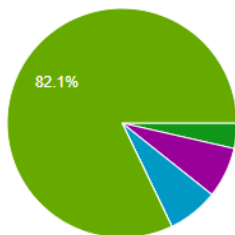
The VADER algorithm outputs text visualizations with positive and negative sentiments in word clouds that highlight words which are more frequently appearing in statements. Word clouds are used in business, politics, entertainment, and in education for concept extraction and

public pulse identification through words that highly appear in the cloud that reflects the context or sentiment of a collection of statements or expressions (Atenstaedt, 2012). Highlighted words are interpreted based on the context of how they were expressed during the personal conversation, hence coming up with the sentiment of the respondents in relation to the research questions raised during the interview.

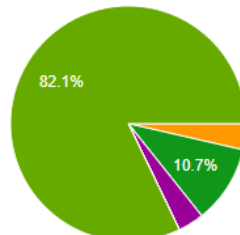
Results and Discussion

A total of 28 participants responded in the survey questionnaires and their responses were duly recorded, transcribed, and translated. Around 72% of the respondents are in the 19-20 years of age and the rest are between 16-18, and 21 years of age. Majority are Filipino students where 72% are male and 28% are female. All of them are using social media where 96.4% of them have multiple accounts. The top 3 social media platforms where they are engaged are Facebook, Instagram, and Twitter. Most of the video contents are viewed from YouTube. The social media usage of the respondents a week before the interview was asked and the results are shown in Figure 2. Highest usage is found upon waking up, between waking-up and before breakfast, between breakfast and lunch, between lunch and dinner, and after dinner and before going to sleep.

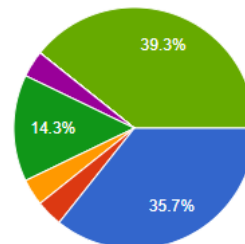
Upon waking-up



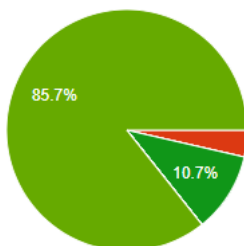
After waking-up and before breakfast



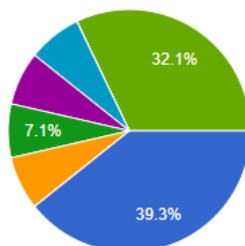
When eating breakfast



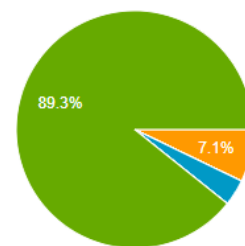
After breakfast and before lunch



When eating lunch



After lunch and before dinner



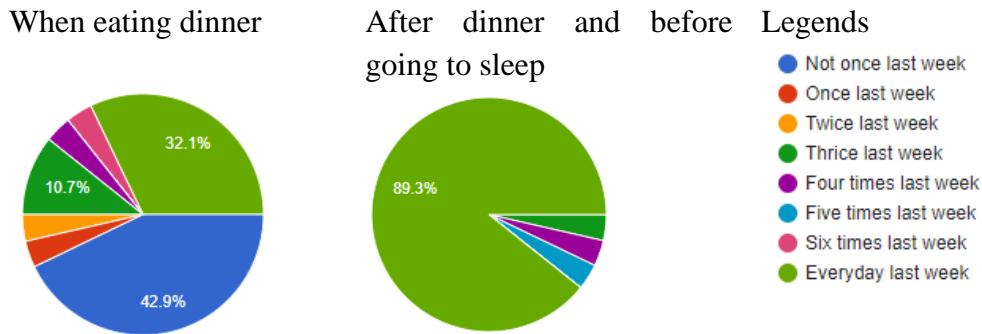


Figure 2: Social media engagement of the respondents during the week considered

The VADER algorithm outputs text visualizations with positive and negative sentiments in word clouds which highlight words that appear more frequent in statements. Figure 3 shows the cloud words for the responses of the participants in RQ1, RQ2, and RQ3, respectively.



(a)



(b)



(c)

Figure 3: Word clouds for (a) RQ1, (b) RQ2, and (c) RQ3 depicting the significant words in the participants’ responses

Table 1 shows the summary of significant words that indicate the positive and negative sentiments of the respondents. It is noticeable that the word “friends” implies positive sentiment while “Facebook” and “post / posting” imply negative sentiment.

Table 1: Summary of positive and negative sentiments

	Positive	Negative
What is social media to you?	<u>friends</u> , feel, share	Social media, time, <u>Facebook</u> , check, <u>post</u>
What are the reasons of social media engagement?	<u>friends</u> , share, feel, entertained	<u>Facebook</u> , feel, <u>post</u>
What are the social media contents that make you engaged?	endorsement, <u>friends</u>	Social media, <u>posting</u> , news, <u>Facebook</u> , classes

Getting virtually connected with friends, and family members keep the mind stress free. Moreover, diet, exercise, and enough sleep support mental health by mitigating anxiety (Sundarasan et al., 2020). Hence, “friends” is a good word to say as it could bring-in positive vibes and sentiments to the young. Friends are considered as allies and this brings in security, companionship, and confidence. Conversely, “Facebook” and “post / posting” are words that implies negative sentiments. The functional correlation between these two words is high since the Facebook social media platform contains a large number of posts which came from different sources around the world. In a study conducted by (Jha et al., 2016) among students in the allied health science fields, almost two-thirds of the respondents admitted that Facebook has had negative impact on their studies as they experienced burned out, disturbed sleep, and headaches due to long exposure and engagement in the platform. The brand name might have brought these concepts to the respondents or might have had experienced the same as with the health allied students that would have caused its sentiment to be negative. Furthermore, as time spent in social networking sites increases, the academic performance of the students is seen to deteriorate (Abdulahi et al., 2014) hence, a clear indication of its negative sentiment.

Conclusion

Thoughts expressed in words and represented by texts are useful in determining the sentiments of a person towards another person, things, events, or an experience. Using NLP, the words are processed and analyzed using a computer algorithm that estimates the sentiments in phrases, sentences, or paragraphs. For mental health care, it is important to guard human thoughts as it relates to human feelings and emotions that lead action which could either be beneficial or not. Word analysis could be a complementary procedure in analyzing a person’s mental health condition and in composing therapeutic procedures that aims to manage mental conditions. Once the words that bring negative sentiments are identified, they can be avoided and not be used in communication to avoid triggers and other unwanted reactions.

References

Abbas, J., Aman, J., Nurunnabi, M., & Bano, S. (2019). The impact of social media on learning behavior for sustainable education: Evidence of students from selected universities in Pakistan. *Sustainability (Switzerland)*, 11(6), 1–23. <https://doi.org/10.3390/su11061683>

- Abdulahi, A., Samadi, B., & Gharleghi, B. (2014). A Study on the Negative Effects of Social Networking Sites Such as Facebook among Asia Pacific University Scholars in Malaysia. *International Journal of Business and Social Science*, 5, 133–145.
- Atenstaedt, R. (2012). Word cloud analysis of the BJGP. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners*, 62(596), 148. <https://doi.org/10.3399/bjgp12X630142>
- Hutto, C. J., & Gilbert, E. (2015). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*.
- Sehgal.P, Kumar.B, Sharma.M, Salameh A.A, Kumar.S, Asha.P (2022), Role of IoT In Transformation Of Marketing: A Quantitative Study Of Opportunities and Challenges, *Webology*, Vol. 18, no.3, pp 1-11
- Kumar, S. (2020). Relevance of Buddhist Philosophy in Modern Management Theory. *Psychology and Education*, Vol. 58, no.2, pp. 2104–2111.
- Roy, V., Shukla, P. K., Gupta, A. K., Goel, V., Shukla, P. K., & Shukla, S. (2021). Taxonomy on EEG Artifacts Removal Methods, Issues, and Healthcare Applications. *Journal of Organizational and End User Computing (JOEUC)*, 33(1), 19-46. <http://doi.org/10.4018/JOEUC.2021010102>
- Shukla Prashant Kumar, Sandhu Jasminder Kaur, Ahirwar Anamika, Ghai Deepika, Maheshwary Priti, Shukla Piyush Kumar (2021). Multiobjective Genetic Algorithm and Convolutional Neural Network Based COVID-19 Identification in Chest X-Ray Images, *Mathematical Problems in Engineering*, vol. 2021, Article ID 7804540, 9 pages. <https://doi.org/10.1155/2021/7804540>
- Jha, R. K., Shah, D. K., Basnet, S., Paudel, K. R., Sah, P., Sah, A. K., & Adhikari, K. (2016). Facebook use and its effects on the life of health science students in a private medical college of Nepal. *BMC Research Notes*, 9(1), 378. <https://doi.org/10.1186/s13104-016-2186-0>
- MathWorks. (2021). vader Sentiment Scores. https://www.mathworks.com/help/textanalytics/ref/vadersentimentscores.html#mw_c3010cf9-4325-457c-8de4-4bde3f13c336
- Moßburger, L., Wende, F., Brinkmann, K., & Schmidt, T. (2020). Exploring Online Depression Forums via Text Mining: A Comparison of Reddit and a Curated Online Forum. *Proceedings of the Fifth Social Media Mining for Health Applications Workshop & Shared Task*, 70–81. <https://www.aclweb.org/anthology/2020.smm4h-1.11>
- Patton, D. U., Frey, W. R., McGregor, K. A., Lee, F.-T., McKeown, K., & Moss, E. (2020). Contextual Analysis of Social Media: The Promise and Challenge of Eliciting Context in Social Media Posts with Natural Language Processing. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 337–342. <https://doi.org/10.1145/3375627.3375841>

- Statista Research Department. (2021). Share of advertising audience of social media in the Philippines as of January 2021, by age and gender. <https://www.statista.com/statistics/1139168/philippines-social-media-advertising-audience-age-and-gender/>
- Sundarasan, S., Chinna, K., Kamaludin, K., & Nurunnabi, M. (2020). Psychological Impact of COVID-19 and Lockdown among University Students in Malaysia : Implications and Policy Recommendations. *International Journal of Environmental Research and Public Health*, 16(6206).
- Tho, C., Heryadi, Y., Lukas, L., & Wibowo, A. (2021). Code-mixed sentiment analysis of Indonesian language and Javanese language using Lexicon based approach Code-mixed sentiment analysis of Indonesian language and Javanese language using Lexicon based approach. *Journal of Physics: Conference Series*, 1869, 8–14. <https://doi.org/10.1088/1742-6596/1869/1/012084>