Inescapable Media: The Role of Social Media Data Mining in Understanding American Adolescent Media Engagement

Cheryl Siyuan Peng

Yew Chung International School of Chongqing, 2 Huxia Street, Yuan Yang Town, Yubei District, Chongqing, 401122, China

Abstract. With social media as the frontier of many American adolescent lives, there seems to be a shift in experiences of the new generation. This paper delves into the impacts of American adolescent media engagement, with a priority of the role of social media data mining in understanding behaviour and trends. This study utilizes innovative data mining techniques such as sentiment analysis, trend detection, and network analysis; it allows insights into how social interactions impact psychological and emotional growth. The discovery of the twofold nature of the media provides opportunities for affirmative individualism and likewise, perils such as cyberbullying. By applying social media data mining, it contributes to an innate understanding of the effects of the media on youth development. The paper highlights the progress of tactics to mitigate the undesirable effects of social media whilst enriching its progressive influences on growth and welfare.

1 Introduction

In the 21st century, a childhood, previously filled with outdoor activities and huddling around a board game, now consists of endless scrolling through videos on media platforms. In 2020, the worldwide COVID-19 pandemic was a significant factor that induced social media viewing. As vaccines and treatments against this epidemic were limited, governments around the globe resorted to prevalent dealings: corporate and education closures as well as movement bans, referred to as “lockdowns” [1]. It restricted the lives of millions of individuals, resulting in a change in how their newly found free time was spent [2]. Let us take TikTok as a starting example. Those teenagers and young people of the new generation, driven by boredom and curiosity, started on TikTok and became so viral that their names became household uses [3]. In addition to the rise in content creators, the number of viewers skyrocketed as adolescents spent hours scrolling through hundreds of videos daily. As of lately, TikTok has 30.8 million daily active users via IOS [4].

This paper will study American adolescents’ engagement in social media (SM) through the lens of social media data mining (SMDM). In addition to that, the paper will delve into the specific and various of techniques that are utilized during the process, evaluating the

Corresponding author: cheryl.peng2021@ycis.com

applications of data mining (DM). Thus, this can help understand some specific aspects of negative or positive trends in the media engagement, considering the perspectives of the youth.

2 Literature Review

2.1 Engagement

Every swipe, double tap, or click that occurs each hour highlights the governing role of SM in the lives of teenagers nowadays. An article by Amy Stornaiuolo studies how SM might impact the maturity of teenagers. As the journal article was a commentary on Boris Zizek's study on digital socialization, Stornaiuolo remarks that there is a need for an alternative method to study interactions [5]. Her ideas are also supported by Borzekowski, who describes a new construct within the abstract of her article: constancy. This term connotes the constant state that both children and adolescents being glued to their screens [6]. Constancy is focused on the impact of the media and can sway the growth of children as well as "future generations" cognition and education, social interactions, emotions, and health [6]. Therefore, constancy may be the alternate aid for the technique to study interactions that Stornaiuolo was trying to put emphasis on within her article.

Continuing, an article written by Martin Lalonde et al. examines three variables – civic involvement, media, and visual arts – and determine if they can reintegrate these within the education of the "at-risk" youth [7]. The general conclusion of the investigation highlights the freedom that adolescents should feel when engaging on SM such as not being swayed by the media culture. When juveniles interact on the media, there is a possibility that they feel this freedom that Lalonde et al. described, however the article "The Impact of Social Media on Society: A Systematic Literature Review," describes some negative aspects that come with engaging on SM. The article "synthesizes" research and discoveries (2016-2024) from more than 30 peer-reviewed articles, case studies, and reports [8]. The closing thoughts of the article emphasizes that there are benefits and challenges to SM, though at sometimes the negatives might outweigh the positives. They put importance on the urgency to promote healthy habits and harmonious attitudes [8].

2.2 Adolescents

The teenage years are quite complex. During this time, their experiences shape their identities, as seen in Victoria Berumen's article, where she breaks down the sociological implications of SM in a positive way. This includes how there are many communities online, and those who find a connection with the "interest, values, and beliefs" feel validated [9]. However, there is a concern of false information circulating in the media, which transitions into Florence Martin et al.'s article; it further delves into how the media can be both fallacious and beneficial. As the study found, a noteworthy number of very young individuals start on social media; hence, there are many apprehensions involving the high probability of cyberbullying and privacy concerns [10].

Moving on, the next two pieces of literature, written by Amber Virden et al. and Janine Senekal et al., both bring light to the associations of teenagers and the media. Yet, it is vital to note that the papers bring alternative viewpoints with their categorical core. Virden et al. are more focused on the emotional and cognitive risks that social communication may bring to teenagers [11]. This includes activities such as sexting, cyberbullying, online hate, etc. [11]. On the other hand, Senekal et al.'s research is more directed at the widespread implications on adolescents. Youth development, mental health, and social interactions were

just a few that were mentioned [12]. Together, both contain and share knowledge about how these adolescents manoeuvre their way through this digital world.

3 Social Media Data Mining

What is social media data mining (SMDM)? To answer that, SMDM is essentially a branch of data mining. Rather than focusing on a broad landscape in the fields such as business and scientific research, SMDM solely focuses on data from SM platforms (i.e. Facebook, TikTok, X, etc.) Evidently, both will identify patterns and trends in data. More specifically, in SMDM, this data is cumulated and then looked through thoroughly to pin down trends [13]. It is said that SMDM searches for hidden patterns that other traditional systems cannot not gather [14]. During the process, many algorithms are utilized to track analytics and then transforms them into readable documents for research or other processes [13]. A few data types that may be collected are the following: reactions, engagement metrics, impressions, and more [15].

The journal article "Social Media – From Social Exchange to Battlefield" mainly focuses on the drawbacks of the media. The first part of the article goes through emblematic moments of history where new, vital SM platforms rose. Bialy puts emphasis on the fact that these much-publicized media platforms have now been altered to a place where individuals speak out about informational warfare in great lengths [14]. However, a particle point raised in this article contributes to this paper's research – data mining, specifically, SMDM. Bialy states that "Social media mining, draws on the different disciplines of computer science, machine learning, social network analysis, statistics, sociology, and many other. [14]" In addition to that, the article also brings to light the concept of big data. Big data can be easily said as datasets that are so enormous and so multifaceted that conventional software, used for data processing, cannot handle them [16]. The article notes that both big data and SMDM are evolving concepts with a bright future and an increased number of possibilities [14].

3.1 Social Media Data Mining Technology

Even though SMDM can be explained as a branch of data mining, the concept itself also has many layers to it. SMDM uses plenty of algorithms and systems when applied to the data. Let's focus on the first part of SMDM technology and three vital parts.

The first to take note of is sentiment analysis; it makes use of natural language processing to assess what emotions are portrayed through on-screen communication [17]. This is done by closely examining the text. It then discovers the tone of the memo; it could be neural, positive, or negative [18]. Sentiment analysis can be used in places such as comments and reviews. This can be important since it gives a clear understanding of what might be needed for a company or service provider [18]. Moving on, another important aspect of SMDM is trend detection. The definition of trend detection is simple to understand. It essentially detects trends which include specific patterns, slight alterations, and any progress in data [19]. From the human eye, we can notice the major changes however, for those miniscule variations in trends, it is nearly impossible to distinguish them; therefore, this method can be used. Clustering, the last method that this paper will focus on, is essentially the process of categorizing data that are like each other [20]. It is commonly used in machine learning but also important in SMDM. Visual representation is important for the delivery of the knowledge gained. In the case of adolescents, it would be specifically aligned with what may be prevalent.

3.2 Application of Social Media Data Mining in Youth Behaviour and Trends

Anastazia Zunic, Padraig Corcoran, Irena Spasic, at the School of Computer Science and Informatics in Cardiff University explores the cognitive representation of adolescents on SM in their research paper. A critical part within their paper is the application and process of SMDM. Sentiment analysis was the first technique that was applied. This provided them with acumens into the expressive statuses of juveniles in which their virtual conducts are prejudiced by their demand for peer endorsement [21]. Moving forth, text mining techniques were applied to recognise subjects of conversations amid pubescents [21]. Lastly, network analysis was applied, which allowed the researchers to find key influencers within communal nets and comprehend the stimuluses of digital exchanges [21]. The tactic made in the paper offered valued visions into youth behaviour but also has wider insinuations for understanding the effect of SM on growth.

From that, it can be understood that for our study, a sentiment analysis model should be created. To start, a sentiment analysis model was sourced from TechVidvan where we plan to alter it to specific data (i.e. youth trends) [22]. To ensure that the model is accurate, a loss function, which is represented by MSE and can be applied to the model.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1}$$

After several trials, the fifth model can be deduced by decreasing the loss function result over and over. This model can then be utilized on any data related to SMDM in youth trends.

4 Ethical Considerations

The importance of ethics has always been undermined. Especially in these situations where data mining, in some perspectives, violate many rules across the board if performed in a malicious intent. However, ethics are also personalized. A person may think that a search for personal data by data mining is not unethical, however another individual may take this as a serious offense. Therefore, the best way to understand how the ethics of data mining and big data is to explore the different viewpoints of established researchers.

The perfect article to start this analysis off is Floridi and Taddeo's article titled, "What is data ethics?" Through this, they develop their take on data ethics and speak about how there needs to be a general outline so that future researchers and analysts can approach their tasks with ease, amplifying their outcomes [23]. In the end, the authors end up developing and suggesting a theoretical skeleton that embodies the main concepts of data ethics. There are three main axes: the ethics of data, algorithms, and practices [23]. They go into detail for each axis, but overall, the framework is more general and is proposed. Furthermore, the book written by Annika Richterich similarly focuses on ethics, however it is mainly for big data. She considers opposing opinions and contributes her stance on the matter. Almost contradicting the earlier article, Richterich believes that the conventional outlines for ethics are not fit for the research of big data [24]. Specifically, chapter three introduces concepts, detailed for big data, such as "group privacy," "weaponized transparency," "emotion contagion," and "data philanthropy" [24].

Overall, some important keynotes that can be taken away from the information recalled. For example, the weaponized transparency that is brought up by Richterich can be used as an axis for an updated ethic framework for researchers. Yet, it is still worth mentioning that ethics is a major personal feat that one needs to learn.

5 Understanding American Adolescence Media Engagement

So, how does social media mining play a role in understanding American adolescent media engagement? Using the techniques introduced and analysed in Section 3.2, a connection can be made. The steps can be seen in Figure 1.

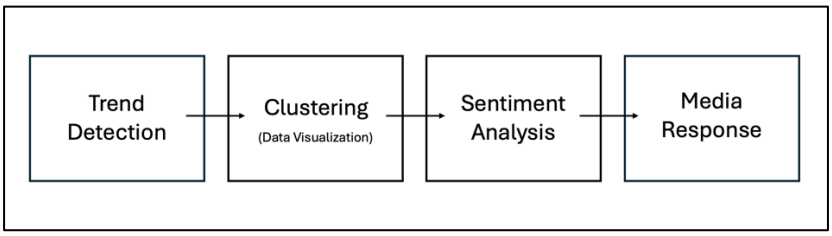

Fig. 1. The steps of utilizing SMDM in understanding adolescent SM engagement.

Going forth with trend detection, patterns and changes can be identified. Trends may include the average daily screen time of teenagers, the average age that children receive their first mobile device, specific activities on media platforms, and more. The data collected from the detection can then be clustered. American adolescents can be divided into many categories: early teen years, late teen years, immigrant parents, ethnicity, family wealth score, etc. From there, additional trend detection can be applied to further analyze the data. Ultimately, based on the results, analysts can understand what to speak out about. Whether it is just a statement, the response of the media can be examined through the sentiment analysis algorithm and therefore, researchers are able to cognize the public by evaluating the emotions of every response.

6 Conclusion

This paper aimed to analyze and investigate American adolescents' engagement in media, especially the impact of SM on their development. Throughout the process, the exploration of the negative and positive impacts was discovered by reviewing several literary pieces that relate. For that matter, readers were able to gain an understanding of how the development of the youth may have been affected. We found that social media plays a dual role in adolescents' lives, providing a platform for positive self-expression and social interaction, but also bringing negative effects such as cyberbullying, anxiety, and loneliness. Among other things, this paper also investigated and researched about the many techniques, algorithms, and methods involved in social media data mining to gain this deeper understanding of the social media activity.

Briefly speaking, there is a need for highlighting the continued increased use of social media from the youth that are not just located in the USA. This paper touched on the fact that childhoods have been altered because of this statistic. As of April 1, 2024, more than five hours daily are spent of media platforms for teenagers [25]. This just proves that more energy needs to be put on restricting the use of these platforms due to the impacts that may be made.

References

1. Robertson, L. (2022, March 8). What We've Learned About So-Called 'Lockdowns' and the COVID-19 Pandemic. Retrieved from FactCheck: https://www.factcheck.org/2022/03/scicheck-what-weve-learned-about-so-called-lockdowns-and-the-covid-19-pandemic/

2. Kessel, P. V., Baronavski, C., Scheller, A., & Smith, A. (2021, March 5). In Their Own Words, Americans Describe the Struggles and Silver Linings of the COVID-19 Pandemic. Retrieved from Pew Research: https://www.pewresearch.org/data-labs/2021/03/05/in-their-own-words-americans-describe-the-struggles-and-silver-linings-of-the-covid-19-pandemic/
3. Kelly, A. (2020, December 11). How Quarantine and TikTok Birthed a New Generation of Young Hollywood. Retrieved from L'OfficieL: https://www.lofficielusa.com/pop-culture/how-quarantine-built-tiktok-celebrities-charli-damelio-addison-rae
4. Shephard, J. (2024, August 07). 25 Essential TikTok Statistics You Need to Know in 2024 . Retrieved from The Social Shephard: https://thesocialshepherd.com/blog/tiktok-statistics
5. Stornaiuolo, A. (2017). Contexts of Digital Socialization: Studying Adolescents' Interactions on Social Network Sites: Commentary on Zizek. Human Development, 60(5), 233-238.
6. Borzekowski, D. L. (2019, December). Constancy (the New Media "C") and Future Generations. Health Education & Behavior, 46(Supplement), 20-29.
7. Lalonde, M., Castro, J. C., & Pariser, D. (2016). Identity Tableaux: Multimodal Contextual Constructions of Adolescent Identity. Visual Arts Research, 38-55.
8. Omar, A. S., & Ondimu , K. O. (2024). The Impact of Social Media on Society: A Systematic Literature Review. The International Journal of Engineering and Science (IJES), 13(6), 96-106.
9. Berumen, V. (2023, August 15). The Impact of Social Media onSelf-Identity Formation: A Psychological and Sociological Perspective. Retrieved from StassOn: https://www.stasson.org/the-impact-of-social-media-on-self-identity-formation-a-psychological-and-sociological-perspective/
10. Martin, F., Wang, C., Petty, T., Wang, W., & Wilkins, P. (2018, January). Middle School Students' Social Media Use. Educational Technology & Society, 21(1), 213-224.
11. Virden, A. L., Trujillo, A., & Predeger, E. (2014). Young Adult Females' Perceptions of High-Risk Social Media Behaviors: A Focus-Group Approach. Journal of Community Health Nursing, 133-144.
12. Senekal, J. S., Groenewal, G. R., Williams, K., Wolfaardt, L., & Jansen, C. (2022, August 22). Social media and adolescent psychosocial development: a systematic review. South African Journal of Psychology, 53(2), 157-171.
13. Ramirez, R. (2020, April 3). Social Media Data Mining: Understanding What It Is and How Businesses Can Use It. Retrieved from University of San Diego, Knauss School of Business: https://www.sandiego.edu/blogs/business/detail.php?_focus=76022
14. Biały, B. (2017). Social Media—From Social Exchange to Battlefield. The Cyber Defense Review, 2, 69-90.
15. Mironov, D. (2024, Feburary 2). From Posts to Patterns: Mastering Social Media Data Mining. Retrieved from Improvado: https://improvado.io/blog/what-is-social-media-data-mining
16. Tiao, S. (2024, March 11). What Is Big Data? Retrieved from Oracle: https://www.oracle.com/big-data/what-is-big-data/

17. Lown, P. (2024, August 21). Social media sentiment analysis: Benefits and guide for 2024. Retrieved from Sprout Social: https://sproutsocial.com/insights/social-media-sentiment-analysis/
18. AWS. (2022). What is Sentiment Analysis? Retrieved from Amazon Web Services: https://aws.amazon.com/what-is/sentiment-analysis/
19. Gaillot, A.-D. (2023, October 27). Trend Detection Definition & Algorithm Guide. Retrieved from Meltwater: https://www.meltwater.com/en/blog/trend-detection
20. Google Developers. (2019). What is clustering? Retrieved from Google Developers: https://developers.google.com/machine-learning/clustering/overview
21. Zunic, A., Corcoran, P., & Spasic, I. (2020, January 28). Sentiment Analysis in Health and Well-Being: Systematic Review. JMIR Med Informatics, 8(1).
22. TechVidvan, Sentiment Analysis using Python [with source code]. (2021, June 14). https://techvidvan.com/tutorials/python-sentiment-analysis/
23. Floridi, L., & Taddeo, M. (2016, December 28). Introduction: What is data ethics? Philosophical Transactions: Mathematical, Physical and Engineering Sciences, 374, 1-5.
24. Richterich, A. (2018). The Big Data Agenda: Data Ethics and Critical Data Studies. University of Westminster Press.
25. DeAngelis, T. (2024, April 1). BY THE NUMBERS. Retrieved from American Psychological Association : https://www.apa.org/monitor/2024/04/teen-social-use-mental-health