International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Leveraging Artificial Intelligence and Machine Learning to Profile Internet Users through Media Content

Abhishek Shukla

Abstract: ***Artificial Intelligence and Machine Learning have progressed enough to be able to profile internet users through data collected via social media content and engagement. In order to create user profiles that offer the greatest accuracy, a system needs to be created to ensure that data is accurately collected and analyzed efficiently. Although such a system would yield great benefits, there needs to be a focus on the ethics of data collecting and the privacy of social media users.***

Keywords: Artificial Intelligence, Machine Learning, Internet Users, Social Media, User Profiles, Data Collection, Data Analysis, Accuracy, Efficiency, Ethics, Privacy

1. Introduction

The development of social media platforms has yielded a vast wealth of content created by the very people who also consume it. These materials include text, audio, and photo/video content that caters to a diverse range of tastes and interests. Aside from the content itself, how users engage with this information also provides very useful information for marketers who wish to understand more about consumers. This digital landscape has given rise to a growing need for efficient and personalized user profiling systems that can help deliver content recommendations and advertisements tailored to individual preferences. In response to this demand, artificial intelligence (AI) and machine learning (ML) technologies have emerged as powerful tools for profiling users based on their interactions with media content. Utilizing AI and ML to profile internet users through the media content they generate is not a new concept. AI and ML have already been harnessed to understand target audiences for products [1]. However, there currently is no industry standard for how data is collected and analyzed, meaning there is no standard for the accuracy of user profiles that are created, and there is limited ethical consideration for the users whose data is collected. This essay explores the development and implications of creating a system for profiling users using media content on the Internet through AI and ML.

2. Methodologies for User Profiling

Developing a system for generating user profiles from media content requires a robust system for collecting and analyzing data. An example of such a system is illustrated in Figure 1, which outlines the process of data creation, logging, extraction, classification, analysis, behavior analysis, to the creation of a user profile [2].

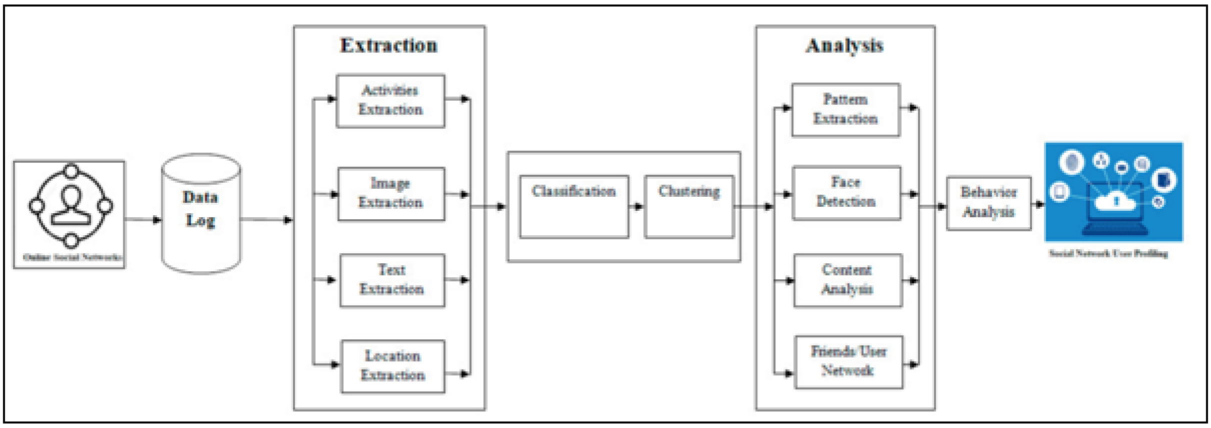

Figure 1: An example of a system that utilizes artificial intelligence and machine learning to general user profiles based on the social media content of internet users [1]

Through this system, social media users interact with various platforms. These interactions (and any content that is created), are collected in a data log [3]. There is a wealth of information sources that can be used to achieve this, across a range of mediums. By nature, most social media platforms today gather information on user-generated content such as uploaded photos, text, or audio, but also interactions such as search queries and links clicked, and IP address [4]. AI and ML algorithms rely on feature extraction to convert raw data into meaningful representations. In the context of media content, this may involve extracting audio features (e.g., spectrograms, pitch, and tempo) or video features (e.g., frame statistics, object recognition, and sentiment analysis) [5]. These features are critical for building models that capture user preferences and behavior.

Once the data has been collected, it is classified and categorized based on similarities [6]. After distinguishing this, data can then be clustered based on similar categorization [1]. This helps to streamline the big data into segments that AI and ML algorithms are capable of analyzing and using for accurately creating user profiles to predict future behaviors.

Once the data has been sorted, it is possible for AI and ML systems to thoroughly analyze it. There are a number of methods or filters of analysis, depending on the type of content that is being analyzed. Firstly, the data will be measured against other data to check for patterns [1]. In images and video, face detection is used to organize data based on the people depicted inside of it [7]. Face detection is used to help identify expressions, and perceptions of agreement or disagreement, which can then be used to imply the users' response to a piece of content [1]. Content analysis serves to understand the purpose or message behind the media, and how this relates to the individual that posted or engaged with it (whether they agreed or disagreed with its sentiment [8]. Data is also used to connect users with other people in their networks, whether these connections are explicit or not [Yang]. This helps AI and ML to create multiple user profiles of a single piece of social media content, or interaction.

After this analysis of a user’s social media activities, this information is then fed into AI and ML algorithms for behavioral analysis. By understanding behaviors exhibited on social media, this technology aims to understand trends and patterns in an attempt to predict future behaviors [10]

Through this systematic process, AI and ML are able to create user profiles for social media content creators and consumers. These user profiles will become more accurate as the algorithms used become more familiar with the data they are collecting and analyzing, but overall these systems can generate accurate and valuable information from a marketing perspective [1]. Based on this system, a user profile is created, and will continue to be refined as more social media interactions by the user are analyzed.

3. Ethical and Privacy Concerns

AI and ML technology offer great opportunities to develop user profiling systems, however, this relatively new technology also poses significant ethical and privacy concerns for social media users. Analyzing audio and video data collected to build user profiles must be done with utmost caution to protect user privacy. A key way to maintain this privacy is the informed consent of data collection [11]. Users should be aware that the social media they create or engage with can be used for the purpose of generating user profiles.

It should be noted that the threat to the safety and security of personal data is not solely caused by AI and ML, but there is evidence to suggest that the terms of service of many social media platforms such as Facebook and YouTube already violate this [12]. Therefore the problem cannot solely be attributed to the collection and analysis of user’s data, but the platform that hosts the data itself should also be considered. Regardless, more transparent consent mechanisms are crucial for maintaining user trust. Robust data encryption and security measures should be in place to safeguard user information from potential breaches - on the social media platforms themselves, as well as within the AI and ML systems that collect the data.

With the current iteration of AI and ML technology, there are widespread biases present in training data, which can filter into big data collected from users, impacting results [13]. AI and ML models used in user profiling may inherit biases present in the training data, potentially leading to unfair recommendations or content filtering. Therefore, mitigating bias and ensuring fairness in user profiling algorithms is a critical ethical consideration [14]. These biases repeat prejudices found in society, including gender biases, sexism, and racism [13]. Therefore it is important to acknowledge that any system designed to create user profiles based on AI and ML analysis of social media content must proactively work to avoid replicating these issues.

4. Conclusion

There is great potential for AI and ML to create user profiles based on social media interaction and content creation. AI and ML technology continue to push the capabilities of generating internet user profiles, and a system designed to collect and analyze data efficiently can generate the most accurate profiles possible, given today's technology. However, addressing privacy concerns, mitigating bias, and ensuring ethical practices are vital to building responsible and trustworthy systems. As the technology continues to develop and become more accessible, any systems that collect and analyze data from social media users must remain balanced between maintaining accuracy with its user personalization, and ensuring all data collected remains private and secure. Future research should focus on refining user profiling techniques, enhancing fairness, and developing innovative methods to protect user privacy in the digital age.

References

[1] J.-A. Choi and K. Lim, “Identifying machine learning techniques for classification of target advertising,” ICT Express, vol. 6, no. 3, pp. 175–180, Sep. 2020, doi: https://doi.org/10.1016/j.icte.2020.04.012.
[2] B. GayathriDevi and V. Pattabiraman, “Towards User Profiling From Multiple Online Social Networks,” Procedia Computer Science, vol. 165, pp. 456–461, 2019, doi: https://doi.org/10.1016/j.procs.2020.01.006.
[3] C. Zachlod, O. Samuel, A. Ochsner, and S. Werthmüller, “Analytics of social media data – State of characteristics and application,” Journal of Business Research, vol. 144, pp. 1064–1076, May 2022, doi: https://doi.org/10.1016/j.jbusres.2022.02.016.
[4] N. Griffioen, M. van Rooij, A. Lichtwarck-Aschoff, and I. Granic, “Toward improved methods in social media research.,” Technology, Mind, and Behavior, vol. 1, no. 1, Jun. 2020, doi: https://doi.org/10.1037/tmb0000005.
[5] O. O. Abayomi-Alli, R. Damaševičius, A. Qazi, M. Adedoyin-Olowe, and S. Misra, “Data Augmentation

and Deep Learning Methods in Sound Classification: A Systematic Review," Electronics, vol. 11, no. 22, p. 3795, Nov. 2022, doi: https://doi.org/10.3390/electronics11223795.

[6] D. Godoy and A. Amandi, "A User Profiling Architecture for Textual-Based Agents," INTELIGENCIA ARTIFICIAL, vol. 7, no. 21, Feb. 2003, doi: https://doi.org/10.4114/ia.v7i21.819.

[7] Z. Stone, T. Zickler, and T. Darrell, "Toward Large-Scale Face Recognition Using Social Network Context," Proceedings of the IEEE, vol. 98, no. 8, pp. 1408–1415, Aug. 2010, doi: https://doi.org/10.1109/jproc.2010.2044551.

[8] H. A. Schwartz and L. H. Ungar, "Data-Driven Content Analysis of Social Media," The ANNALS of the American Academy of Political and Social Science, vol. 659, no. 1, pp. 78–94, Apr. 2015, doi: https://doi.org/10.1177/0002716215569197.

[9] C. C. Yang, X. Tang, Q. Dai, H. Yang, and L. Jiang, "Identifying Implicit and Explicit Relationships Through User Activities in Social Media," International Journal of Electronic Commerce, vol. 18, no. 2, pp. 73–96, Dec. 2013, doi: https://doi.org/10.2753/jec1086-4415180203.

[10] A. Singh, M. N. Halgamuge, and B. Moses, "An Analysis of Demographic and Behavior Trends Using Social Media: Facebook, Twitter, and Instagram," Social Network Analytics, pp. 87–108, 2019, doi: https://doi.org/10.1016/B978-0-12-815458-8.00005-0.

[11] B. Custers, S. van der Hof, and B. Schermer, "Privacy Expectations of Social Media Users: The Role of Informed Consent in Privacy Policies," Policy & Internet, vol. 6, no. 3, pp. 268–295, Sep. 2014, doi: https://doi.org/10.1002/1944-2866.poi366.

[12] M. L. Stasi, "Social media platforms and content exposure: How to restore users' control," Competition and Regulation in Network Industries, vol. 20, no. 1, pp. 86–110, Mar. 2019, doi: https://doi.org/10.1177/1783591719847545.

[13] S. Akter, Y. K. Dwivedi, S. Sajib, K. Biswas, R. J. Bandara, and K. Michael, "Algorithmic bias in machine learning-based marketing models," Journal of Business Research, vol. 144, pp. 201–216, May 2022, doi: https://doi.org/10.1016/j.jbusres.2022.01.083.

[14] A. Tsamados et al., "The Ethics of Algorithms: Key Problems and Solutions," AI & Society, vol. 37, no. 1, pp. 215–230, Feb. 2021, doi: https://doi.org/10.1007/s00146-021-01154-8.