



## Exploring the Role of Social Media in Education through Web Mining: Insights into Student Engagement and Behavioral Trends

Kalokhe Anil Sopan<sup>1</sup> Phadtare Siddhi Dnyaneshwar<sup>2</sup> Kale Samiksha Santaram<sup>3</sup>  
& Pawar Mahesh Dattatray<sup>4</sup>

<sup>1</sup>Research Scholar, Department of Computer Science, Shivaji University, Kolhapur

<sup>2</sup>Student, Department of BBA(CA), Vidya Pratishthan's Arts Science and Commerce College,  
Baramati, Pune(MH), India.

<sup>3</sup>Student, Department of BBA(CA), Vidya Pratishthan's Arts Science and Commerce College,  
Baramati, Pune(MH), India

<sup>4</sup>Assiatant Professor, Department of BBA(CA), Vidya Pratishthan's Arts Science and  
Commerce College, Baramati, Pune(MH), India.

Corresponding Author – Kalokhe Anil Sopan

DOI - 10.5281/zenodo.15501647

### Abstract:

*This research paper explores the application of web mining techniques to analyze how students engage with social media platforms. These studies demonstrate that online communities are valuable data sources which provide rich, longitudinal data that would otherwise be difficult, if not impossible to access. Using a combination of data mining, sentiment analysis, and network analysis, this study investigates patterns in students' social media usage, focusing on platforms such as Facebook, Instagram, and Twitter. In this paper, we developed a workflow to integrate both qualitative analysis and large-scale data mining techniques. Additionally, the paper examines how student interactions on social media reflect their social networks, interests, and attitudes toward education. By leveraging web mining techniques, the study provides insights into the evolving role of social media in students' lives, offering valuable recommendations for educators and stakeholders to enhance student engagement and learning experiences in the digital age.*

**Keywords:** *Crawler, Data Mining, Digital Footprint, Online Interaction, Social Media Usage, Web Information Search.*

### Introduction:

The research paper, exploring the Role of Social Media in Education through Web Mining, investigates how students engage with social media platforms and how their interactions reflect behavioral trends and academic interests. Utilizing advanced technologies to efficiently audit and monitor online content, swiftly detect harmful information on the internet, and prevent the spread of negative digital culture has become a critical scientific and technological challenge [1]. By employing

web mining techniques, sentiment analysis, and network analysis, the study uncovers patterns in students' usage of platforms like Facebook, Instagram, and Twitter. Social media platforms like Twitter, Facebook, and YouTube offer valuable spaces for students to share their experiences, express emotions and stress, and seek social support. On these platforms, students engage in discussions and share their daily experiences in an informal and casual way [2]. The research also addresses ethical considerations, data

privacy, and the implications of digital footprints in education.

The Blogosphere is a subset of the Web and shares most of its general characteristics. However, it differs in certain aspects that influence how it can be modeled and how the model can be used for information extraction [3]. Social media has transformed the way students engage with educational content, communicate with peers, and participate in academic discussions. Recommender systems play a crucial role in the digital landscape, assisting users in filtering through extensive information by offering personalized suggestions tailored to their interests [4]. This paper seeks to enhance social network scholars' efforts to understand the conditions for the responsible dissemination of research findings [5]. While online information can help bridge the knowledge gap between clinicians and patients, it may also be seen as a challenge to a doctor's expertise. Advice obtained from online forums may or may not be accurate, and healthcare professionals may encounter patients who are either better informed than before or have incorrect information [6]. The increasing integration of digital platforms into students' daily routines provides a wealth of data that can be analyzed to gain insights into their behavioral patterns and learning preferences. The evolution of database technology has led to the emergence of Data Mining as a multidisciplinary field focused on data analysis. As a result, Data Mining involves the process of extracting valuable information from large datasets [7]. It highlights how social media serves as a rich data source for understanding student engagement and offers recommendations for educators and stakeholders to enhance learning experiences. Social media are online platforms that enable users to share, exchange information, and interact with content and comments within virtual communities and networks [8]. The study

aims to provide actionable recommendations for educators and policymakers to enhance learning experiences through social media while mitigating risks associated with excessive usage and data privacy breaches. On digital platforms, value is often generated by communities of volunteers who freely contribute their knowledge and skills [9]. Data mining refers to the process of extracting relevant data from large databases. It involves uncovering hidden information within data repositories. In this context, utilizing machine learning to aid in data analysis and pattern recognition is a standard practice [10].

### Literature Review:

Hongjian Guo [1] discussed crawler technology to gather and extract resource information from the World Wide Web, integrating the entire web into a data warehouse for comprehensive web content sampling, auditing, and monitoring. In the subsequent processing stage, techniques such as online analytical processing (OLAP) and data mining, which include web language analysis and logical reasoning, are employed to analyze the data stored in the warehouse.

Xin Chen [2] discussed that this study offers valuable insights for researchers in learning analytics, educational data mining, and learning technologies. It presents a workflow for analyzing social media data for educational purposes, addressing the key limitations of both manual qualitative analysis and large-scale computational analysis of user-generated text.

Jyoti Kadadevarmath [3] observed that social media platforms have become increasingly important on the Web and now make up a large portion of new content. Despite their differences, all types of social media share complex network structures that offer metadata and context, which can be useful for extracting information from their

content. We have presented some preliminary findings from our ongoing work, which focuses on extracting and utilizing this structural information.

### Research and Methodology:

Web mining, when applied to studying how students use social media, involves extracting and analyzing data from various platforms to identify patterns, behaviors, and trends. The research begins by defining objectives and hypotheses, such as understanding students' social media usage frequency, platform preferences, and its impact on academic performance, mental health, or social interactions. A common hypothesis might be that students who spend more than three hours daily on social media have lower academic performance.

Data collection is a crucial step and can be conducted through multiple methods, including web scraping, where automated tools extract public posts, comments, and metadata from platforms like Facebook, Instagram, or Twitter. Another approach involves using social media APIs that provide structured access to user data while adhering to privacy policies. Surveys and questionnaires can also be employed to gather self-reported data on students' social media habits. In some cases, data logging tools may be used to track platform usage by monitoring browsing activity or recording engagement metrics.

Once the data is gathered, it undergoes cleaning and preprocessing to remove irrelevant information, handles missing values, and transform text into an analyzable format using techniques like tokenization, stemming, and lemmatization. After preprocessing, data analysis is performed using various methods. Sentiment analysis helps determine whether students' social media posts reflect positive or negative attitudes toward education. Social network analysis examines how students interact, form connections, and influence

each other. Topic modeling identifies the main subjects of discussions, such as academic topics or mental health. Behavioral analysis focuses on engagement patterns, such as how often students like, comment, and share content, and how these behaviors correlate with academic outcomes.

Ethical considerations play a significant role in such research. Ensuring privacy by anonymizing data, obtaining informed consent when surveys are used, and maintaining data integrity are essential to protecting students' rights and ensuring accurate representation. Finally, the study concludes by summarizing key findings and offering recommendations. These may include suggestions for educational institutions on incorporating social media into learning, guidelines for students on balancing social media usage with academics, and policy recommendations for managing screen time effectively. By following this structured approach, web mining research can provide valuable insights into students' social media behavior and its implications for education.

### Result and Discussion:

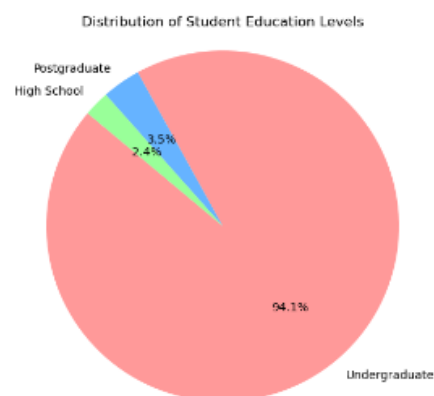


Figure1: The result of distribution of Student Education Levels

After distributing the questionnaires to students of BBA (CA) Department, the first question was that, what is current Education Level you are pursuing? With three answer options (High School,

Undergraduate and Postgraduate) the following results were obtained.

The pie chart shows us student education levels of students. The majority, 94.1%, are undergraduate students, while 3.5% are postgraduate students. Only 2.4% of students are in high school. This indicates that social media is most widely used in undergraduate students.

The next question is what is gender of students? With the answer options male, female and trans. The results obtained from the respondents were as follows:

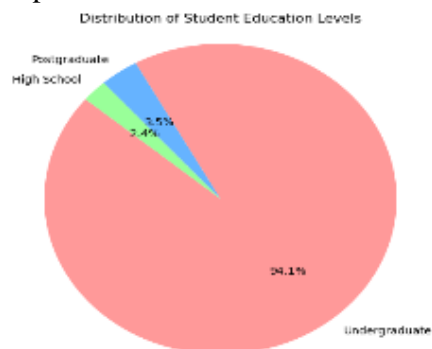


Figure 2: The result of Gender Distribution

This pie chart shows the gender distribution of social media users. The majority, 65.9%, are female students, while the remaining 34.1% are male students. Overall, female students are more active on social media compared to male students.

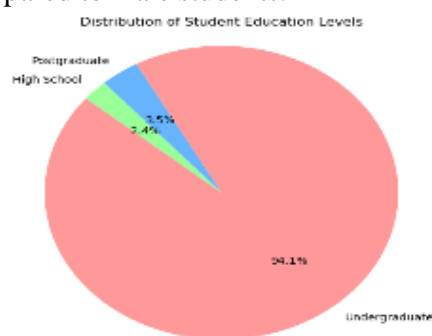


Figure3: The result of Time spent on Social Media

The next question is how much time do you spend on social media platforms on average each day? With the answer options less than 1 hour, 1-2 hours, 3-4 hours, more than 4

hours. The results obtained from the respondents were as follows:

This pie chart shows the time students spend on social media. The largest group, 54.1%, spends 1-2 hours on social media. 3-4 hours comes next at 27.1%, followed by 17.6% who spend less than 1 hour. The smallest group, 1.2%, spends more than 4 hours. Overall, 1-2 hours is the most common duration spent on social media, while more than 4 hours is the least. The next question is do you think social media companies collect data related to your academic behavior (e.g., interests, academic topics you engage with)? With the answer options yes, no and not sure. The results obtained from the respondents were as follows:

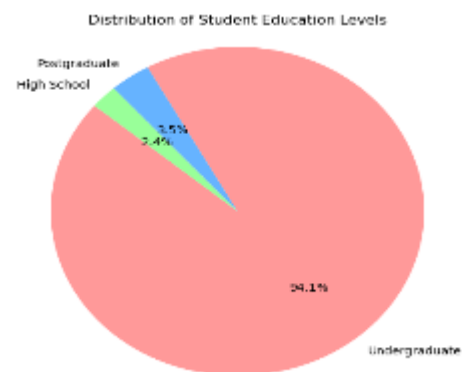


Figure 4: The result of Awareness of Data Collection

This pie chart shows data collected by the companies of students. The majority 77.6% of students say “yes”, indicating they are aware of it, while 16.5% students are unsure whether companies collect their data or not. The remaining 5.9% believe that no data is collected. Overall, most of the students are aware of companies collecting data related to you academic behavior.

The next question is do you think social media will play a more significant role in education in the future? With the answer options yes, no and not sure. The results obtained from the respondents were as follows:

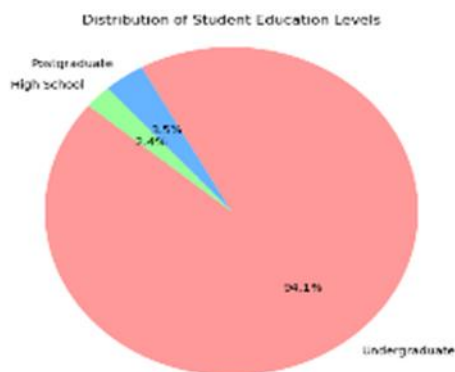


Figure 5: The result of Future role of Social media in Education

The pie chart shows us opinions on the future role of social media in education. The majority, 80.0%, believe that social media will play a role in education, while 15.3% are uncertain about its impact. Small portion, 4.7%, do not see social media as having a role in education. The data suggests that most respondents recognize the potential influence of social media in the educational sector, with only a few expressing doubt or disagreement.

The next question is would you be interested in using a social media platform specifically designed for educational purposes? With the answer options yes, no, maybe. The results obtained from the respondents were as follows:

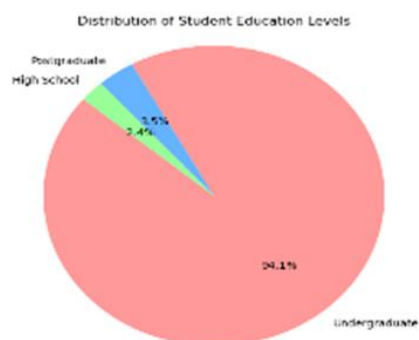


Figure 6: The result of Interest in Educational Social Media Platforms

The pie chart titled Interest in Educational Social Media Platforms represents respondents' willingness to engage with educational content on social media. The chart is divided into three categories: Yes (82.4%) – A large majority of respondents express interest in

educational social media platforms. Maybe (15.3%) – A smaller portion of respondents are uncertain about their interest. No (2.4%) – Only a minimal number of respondents are not interested in such platforms.

The next question is do you prefer learning academic content through videos (e.g., YouTube, TikTok tutorials) or reading articles/posts? With the answer options videos, articles/posts, both equally, neither. The results obtained from the respondents were as follows:

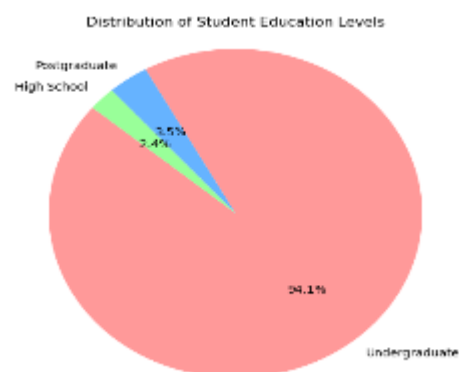


Figure 7: The result of Preferred Learning Format

The pie chart titled Preferred Learning Format illustrates how respondents prefer to consume academic content. The chart is divided into four categories:

Videos (49.4%) – Nearly half of the respondents prefer learning through video-based content, such as YouTube tutorials or educational videos on social media. Both equally (40.0%) – A significant portion of respondents find both videos and written articles/posts equally useful for learning. Articles/Posts (7.1%) – A smaller group prefers reading educational articles or social media posts for academic learning. Neither (3.5%) – A minor percentage of respondents do not prefer either format for learning.

### Conclusion:

This study examines how students use social media and what they think about its role in education. It finds that social media is an important part of students' daily routines, with most spending 1-2 hours per



day on these platforms. Many students are aware that social media companies track their academic interests, showing the need for better digital literacy and privacy awareness. Additionally, 80% of students believe social media will be essential in education in the future. Using web mining techniques, this research identifies trends in how students use social media and interact with content. It suggests that schools and educators can use social media for learning, discussions, and collaboration, but also stresses the need for responsible usage. Issues like privacy risks and excessive screen time should be managed through proper policies and education.

In the future, further research can look at how social media affects academic performance, mental health, and social skills over time. Technologies like artificial intelligence and machine learning can help make learning more personalized. As digital tools continue to evolve, social media's role in education will likely grow, requiring a balanced approach to use its benefits while reducing risks.

### References:

1. Hongjian Guo, "Research on Web Data Mining Based on Topic Crawler," *Journal of Web Engineering*, 24 June 2021, Vol-20, No-4, 1193–1206, DOI: 10.13052/jwe1540-9589.20411.
2. Xin Chen, Mihaela Vorvoreanu, Krishna Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences," *IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES*, July-September 2014, Vol-7, No-3, DOI: 10.1109/TLT.2013.2296520.
3. Jyoti Kadadevarmath, "Study on Social Media Using Web Mining," *Journal of Computational Analysis and Applications*, 20 December 2022, VOL. 30, NO. 2, 355-359.
4. Supriya Saxena, Bharat Bhushan, "Transformers (BERT) Based Framework for Web Recommendations Using Sentiment-Enriched Web Data," *Journal of Information Systems Engineering and Management*, 12 January 2025, e-ISSN: 2468-4376, DOI: <https://www.jisem-journal.com/>.
5. Paola Tubaro, "Whose results are these anyway? Reciprocity and the ethics of giving back after social network research," *Elsevier B.V. This is an open access article under the CC BY-NC-ND license*, 13 November 2019, 0378-8733, DOI: <https://doi.org/10.1016/j.socnet.2019.10.003>.
6. Jasmina Rueger, Wilfred Dolfmsma, Rick Aalbers, "Perception of peer advice in online health communities: Access to lay expertise," *Elsevier Ltd. This is an open access article under the CC BY license*, 2 September 2020, 0277-9536, DOI: <https://doi.org/10.1016/j.socscimed.2020.113117>.
7. Verma Gulhati Manjula, Yumnam Jayanta Singh, "Web Mining for Social Network Analysis: A Review, Direction and Future Vision," *Journal of Engineering Technology*, ISSN: 2348-7305, Volume 4(1), 2016.
8. Kalokhe Anil Sopan, Pawar Mahesh Dattatray, "A Study On Data Mining Techniques In Social Media Data: A Review," *Aayushi International Interdisciplinary Research Journal (AIIRJ)*, Special Issue-49, 480-485, 13 November 2019.
9. Johannes Loh, Tobias Kretschmer, "Online communities on competing platforms: Evidence from game wikis," *Strategic Management Journal published by John Wiley & Sons Ltd*, 28 July 2022, DOI: 10.1002/smj.3442.
10. S. Senthamaraiselvi, K. Meenakshi Sundaram, "Predicting Online Social Network Student Performance Using Enhanced Random Bayesian Algorithm," *International Journal of Intelligent Systems And Applications In Engineering*, 29 November 2023, IJISAE, Vol-12, No-2, 384–390.