Enhancing Visibility on Social Media with Categorization Machine Learning Analysis

Haitham S. Hasan[1,*]

[1]Business Information Technology Department, Business Informatics College, University of Information Technology and Communications, Baghdad, Iraq

Email: Haitham@uoitc.edu.iq

Abstract

A considerable number of individuals concentrate their engagement on social media sites, notably Instagram, YouTube, and Facebook, where they may adeptly exploit their popularity. A considerable volume of research studies has been conducted across diverse social networks to examine user profiles and their relationship with popularity. The primary emphasis of research concerning social media has centered on theme analysis, encompassing domains such as health, creativity, and awareness. This study utilizes K-means clustering to classify social media articles and determine the characteristics that contribute to their popularity. Gathered data from publications by international influencers during an eight-month duration. Producing roughly 161 posts daily and around 1092 posts monthly seeks to improve metrics including views, likes, and dislikes on social media. This strategy is designed to facilitate the growth of the platform's popularity, thereby maximizing visibility and outreach. The analysis focused on three factors: virility, appeal, and publicity rates. Classified the posts into five fundamental groups: casual, quality, number, support, and leader. The study yielded important findings on optimal publication timing, ideal video length, follower metrics, biography and caption lengths, and hashtag utilization. Researchers found some interesting things that will help people who use social media and brand owners make better marketing plans.

Keywords: Data Science Techniques; Data Pattern Identification; K-mean cluster; Statistics; Machine Learning

1. Introduction

As time has progressed, Social networking is quickly becoming an important part of our daily lives. In addition to networking, it conveys thoughts, information, and insights. As an additional benefit, it enables a large number of people to acquire financial resources by highlighting their reputation as influential persons [1] , on the other hand, users in the corporate world utilize social media for marketing objectives [2]. As of January 2022, YouTube and Facebook boasts approximately one million active members, positioning it as the sixth most popular social network. Because it has a strong, focus on video content, Instagram stands out among other social media networks. Facebook and Instagram use not only to advocate for opinions but also to share information through videos [3]. Social media popularity stems from Due to its user-friendliness, difficulties, and duets, and humor, and its algorithm that offers personalized suggestions to users.

It is possible to generate revenue from Instagram through a variety of means, including video commercials and collaborations with corporations. The primary objective is to increase the number of views, likes, and dislike on YouTube in order to boost the audience for the platform. Examining user attributes and postings on YouTube can determine the factors that determine popularity on the platform [4]. Past studies on social network popularity primarily fall into two categories: user profiles and popularity analysis. Researchers have conducted a limited amount of research on YouTube, while Twitter's popularity has surged due to its API accessibility. User profiling involves the methodical collection and classification of users based on a variety of variables to provide consumers with individualized suggestions [5]. It serves various purposes, such as identifying fraudulent users, facilitating the display of advertisements, and enabling the customization of content. Within the scope of the popularity research, the objective is to identify the elements that influence the prominence of a user [6] This study does not evaluate user profiles in the same way that other studies do; rather, it examines posts. Despite extensive research

DOI: https://doi.org/10.54216/FPA.180203
Received: August 08, 2024 Revised: November 01, 2024 Accepted: January 11, 2025

on YouTube and comprehensive analyses of Sorting and user information in different social media sites, few research focus specifically on the classification and analysis of posts generated on YouTube. The emphasis on short videos and associated features sets YouTube apart from other social networking platforms. So, it's not like other social media sites in terms of content, subscribers, or network [7]. For content creators who are interested in producing brief instructional films, YouTube offers a platform and tools that are of excellent utilization. Learn on YouTube publishes hundreds of videos every day, garnering 72 billion views. The incorporation of music, audio, and textual elements within the video poses a challenge in extracting knowledge domains from the instructional information included in posts. YouTube distinguishes itself from other social networks by its emphasis on both short and long-form videos, along with accompanying features designed to enhance user engagement. These aspects, which encompass posts, users, and the network, provide distinguishing qualities that set this platform apart from others [6]. The primary objective is to enhance the metrics of views, likes, and dislikes on YouTube to facilitate the growth of the platform's popularity. Analysing the characteristics of the shared material and the platform's users is essential for identifying the factors that influence popularity [9]. The objective is to classify social media messages and identify their appeal. We divided postings into informal, quality, volume, supportive, and influencer categories. The study found the best times, days to publish videos, the best video duration, the number of followers, the length of bios, captions, this study will help social media users, and brand owners create marketing strategy. This study categorizes social posts and examines their popularity [10]. The goal of this study is to look at how timely posts affect how popular they are. Does popularity affect post quality? How are they different? Reviewing literature, describing methodologies, analysing data, summarizing findings after categorizing data, drawing conclusions, and recommending future research are all part of the project [11].

2. A Study of Important Literature

This section provides an analysis of the existing research on popularity evaluation, user profiles inside social networks, and other studies that are relevant to social media. According to the findings of a research, study [12] the video posted on some social media such as YouTube that had comedy and humanization received the highest level of interaction [13]. The researchers use a small dataset consisting of 892 videos to classify the material into several categories, such as collections, events, instructional films, miscellaneous, services, and places [14]. Research shows that influencer endorsements generate more traffic than in-feed advertising [15] and people perceive influencer marketing as more authentic than other marketing tactics [16,17] Research on the combination of social media and brand ads has been extensive. There is a correlation between the beauty and credibility of an Instagram influencer and the likelihood that they would engage in impulsive purchasing behaviour as a result of their endorsements [18] . The results of a study that was done on 3,670 the social network brand pages. Revealed that an increase in the frequency of posting is associated with an increase in the degree of engagement (likes and comments), but this impact begins to diminish with time [19]. Online shopping and social networking sites have user profiles. Provide a multitude of benefits, and an in-depth investigation of the importance of social media platforms may reveal valuable information for businesses and influencers alike [20]. This tool specifically designs its product catalogues to cater to a wide range of customer demographics, especially on sites for shopping online. Social media platforms make it easier to distribute news in a targeted manner and to advocate for political candidates. The literature study shown in [21] reveals that user profiling may be achieved by the use of machine learning techniques, more especially K-Means clustering, which is well known for its ease of use and effectiveness in producing balanced results. The theme research includes a qualitative analysis of specific content categories. Some examples of these categories are perspectives on electronic cigarettes [22], cultural differences between TikTok and Douyin [23], and academic libraries [24]. TikTok is primarily concerned with spreading awareness, particularly with regard to what the COVID-19 outbreak did to health and the difficulties associated with cyber bullying. Qualitative research [25,26] demonstrated that YouTube serves as an excellent medium for the growth and distribution of creative abilities through the use of short films[27]. Rather than being based on a user-centric design philosophy [28].

3. Methodology

3.1 Identifying Features of the Data

Table 1 presents the optimal posting times for social media platforms, including YouTube and Facebook. Table 2 delineates the attributes obtained from the gathered data. The attributes are classified into groups: post attributes, user characteristics, and statistics. Characteristics of posts are essential for comprehending the subject matter, upload duration, and video resolution. The timestamp of the post was taken into account while assessing the user attributes. The metrics were classified from the data gathered during the second phase, which commenced thirty days following the report's release.

Posting on Social Media is most effective between Tuesday and Thursday between the hours of 10 and 11 a.m. Every Monday at 6, 10, and 10 o'clock in the evening.

• On Tuesdays, two, four, and nine in the mornings when: Wednesday, 7, 8, and 11 a.m.

DOI: https://doi.org/10.54216/FPA.180203
Received: August 08, 2024 Revised: November 01, 2024 Accepted: January 11, 2025

There will be three sessions on Thursday: at 9, 12, and 7 o'clock. For example, on Fridays at 5, 1, and 3 p.m. • On Saturdays, at 8 p.m., 7 p.m., and 11 a.m. Every Sunday at 7, 8, and 4 p.m.

Table 1: Best time to post on Tik Tok

Time	Monday	Tuesday	Wednesday	Thursday	Friday	Sunday
00 to 3 AM		X				
3 to 6 AM		X			X	
6 to 9 Am	X		X	X		X
9 to 12 PM	X	X		X		
12 to 3 PM					X	
3 to 6 PM					X	X
6 to 9 PM				X		
9 to 00 PM	X		X			

Table 2: Extracted Characteristics

Extracted Characteristics	
Like count	LC
Video Time & Size	VTS
Day of Week	DW
Caption Character	CAC
Followers Count	FC
Dislikes Count	DLC
Shares Count	SHC
Comments Count	CC
Save count	SAC

3.2 Analysis of Unconfirmed Information and Data

This part of the article is meant to offer a better knowledge of the elements that affect the outcome of the situation. Individual popularity on social media by conducting an analysis of a variety of characteristics. In addition to providing descriptive information, this part examines the date and time of the publication.

This section provides a summary of the statistical measures.

Table 3 presents the descriptive statistics for each attribute. Most characteristics have positive skewness, meaning their values cluster closer to the minimum than the maximum. For descriptive statistics, we can numerically study four categorical features. These characteristics are LC, FC, CC, and DW. Day and Dow's skewness is steadily approaching zero, signifying a balanced distribution of posting days. The DW characteristic exhibits negative skewness, suggesting that posting times tend to cluster around midnight (23:00). The gathered the information from Iraqi influencers, who we believe to be located in the GMT+3 time zone. According to the findings of the

DOI: https://doi.org/10.54216/FPA.180203
Received: August 08, 2024 Revised: November 01, 2024 Accepted: January 11, 2025

research, users post captions that are an average of 77.6 characters long. The figure of 21.88% falls in the middle of the range, which spans from 0.91% to 3.77,62%. This shows that TikTok interaction is always changing.

Table 3: Statistics of the Extracted Characteristics

Extracted Characteristics	K-Mean	standard deviation	Variance Skewness	sum and error	Range margin of a given data set
LC (like Counts)	77.6	55,900	0.89	0.091% - 3.77,62%	21.88
FC (Follower Counts)	56,77	44,88	0.55	0.03% -2.53%	2.90
CC (Comment Counts)	44,9	33,87	1.77	0.001% - 0.01%	3.55
SHC (Share Counts)	32,98	22,76	0.21	0.66% -1.02%	0.23
SAC (Save Counts)	24.53	30.43	1.55	0.22% - 2.43%	3.77

3.3 Connections between the Characteristics

This section investigates the relationship between numerical attributes, emphasizing the identification of variables that influence the metrics. In addition to providing a description of the Pearson correlation, also includes the p-values for the three scenarios. Additionally, we have eliminated raw metrics from the table, as the major metrics currently incorporate these values. Even though the R-value usually shows a weak relationship (≤0.4), some characteristics consistently show a strong association (at a p-value of ≤0.001) with the main metrics, such as Hashtag attraction, and Like. Although the link between the characteristic and the measure is not linear, The R-value shows how strong the linear connection is. The connections between the qualities and the metrics are not particularly strong, as expected; no single trait can meaningfully explain the complexity of popularity in social networks. The correlation coefficients for the frequency and amount of postings (DW and LC) are 0.1 and 0.2.8, respectively, indicating that there is a negative link between exposure (followers) and exposure. The volume post category exhibited a low exposure rate, with an average of 6.17%, and a high frequency of postings. Since this is the case, users are required to maintain a sufficient frequency of postings. A previous study found a negative association (R=-0.10) between exposure and the number of followers (FC), suggesting that those with a larger audience had a lower engagement rate [26]. Additionally, a higher follower count (FC) are both associated with a lower exposure (R = -0.24). Therefore, we encourage users to maintain authenticity in their profiles by refraining from creating amassing an excessive number of followers. According to the information, provided in Section the best metrics for maximum exposure for cluster C2 (DW posts) are 33.78 (FC) and 61.47 characters. The R-value of each characteristic provides an indication of the correlation, and the starting number represents this value. When the p-value is less than 0.03, the majority of the data are significant, and the peso code indicates which values to exclude.

Python sklearn for K-means clustering. Container grouping. using zip(x, y) as data and inertias as a list: for each i in range(4): calculated using the kernel with n clusters set to i data.fit by kmeans add inertias to kmeans 'Elbow Method', plt.title, prt.xlabel('Cluster count'), and inertias are plotted using the 'Elbow Method' function. Here: • x is the feature value, y is the mean, and i is the extracted characteristic standard deviation; plt.ylabel('Inertia') plt.show().

K-means is a prevalent clustering method utilized for profiling applications. [22]. The famous "Big Five" personality traits gave us the idea for the number of groups, K=5. [13], K=5 was also utilized in numerous user-profile research [8]. The common method to choose the ideal K for K-means is the elbow method; however, this did not produce a noticeable elbow for this study's data, as seen in Figure 1, which means K-Means produced a good separation between clusters regardless of the chosen K

DOI: https://doi.org/10.54216/FPA.180203
Received: August 08, 2024 Revised: November 01, 2024 Accepted: January 11, 2025

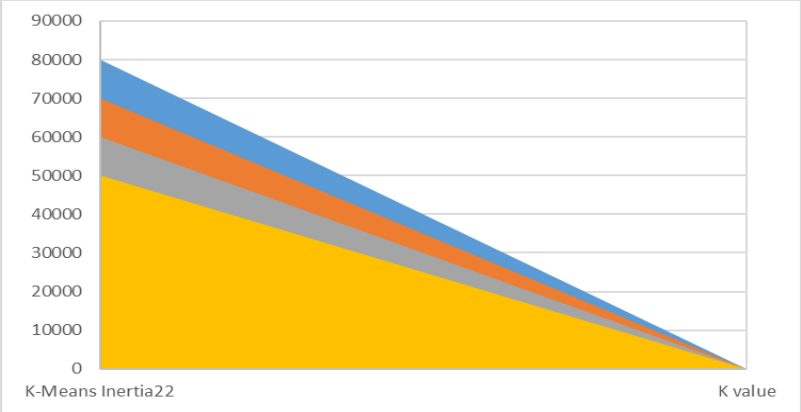

Figure 1. K-Means Result with Various K Values Compared to K-Means Inertia

The study findings categorize posts on social media into five distinct categories: like counts, with a minimal strategic emphasis; share count, with a substantial strategic emphasis; comments counts, dislike counts, follower counts, emphasizing referencing or liking what other people post; and vital posts, which are regular posts from users who have a lot of followers. The majority of the time, casual articles are able to get the highest possible prominence. The postings display an average of 3369 posts from start to finish. The initial distribution of content on social media is illustrated in Figure 2.

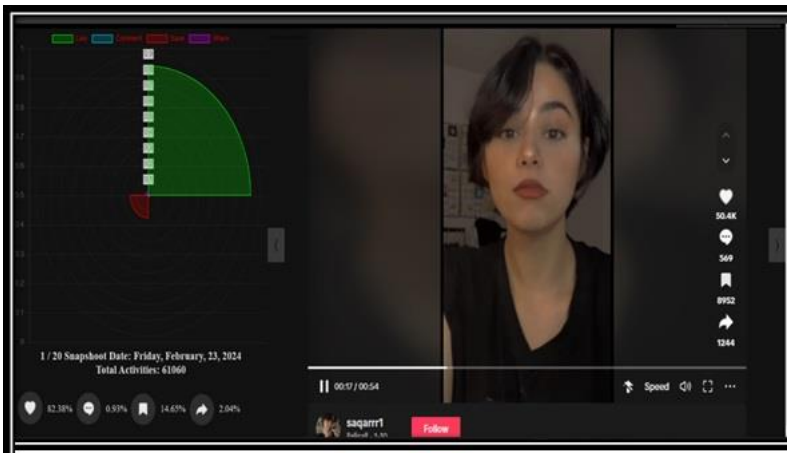

Figure 2. The initial distribution of content on Social media

Figure 2 is a pictorial representation of the major user interface that this study will provide. This presentation includes an aggregated analysis of elements such as shares, follower, comments, like and save. Additionally, this show incorporates social media postings sourced from the program. In Figure 3 is a pictorial representation of the major user interface that this study will provide. This interface includes any social media posts gathered through the application, along with the gathering and analysis of attributes such as shares, save, comments, follower and likes and dislike.

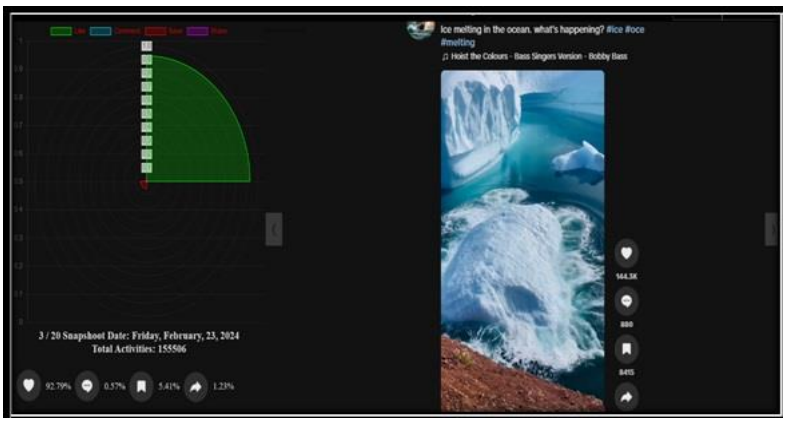

Figure 3. is a pictorial representation of the major user interface

DOI: https://doi.org/10.54216/FPA.180203
Received: August 08, 2024 Revised: November 01, 2024 Accepted: January 11, 2025

In Figure 4 presents the primary user interface that this study will utilize. This presentation includes an aggregate and analysis of quality metrics such as shares. Additionally, this show incorporates postings sourced from the social media program.

Figure 4. presents the primary user interface

The color green represents the number of likes associated with each individual post. The color blue represents the number of comments associated with each post. The color blue represents this quantity. Users use the color red to indicate the actions they have taken to save each individual article. The color purple is a representation of the interactions that individuals utilize to share their content with one another.

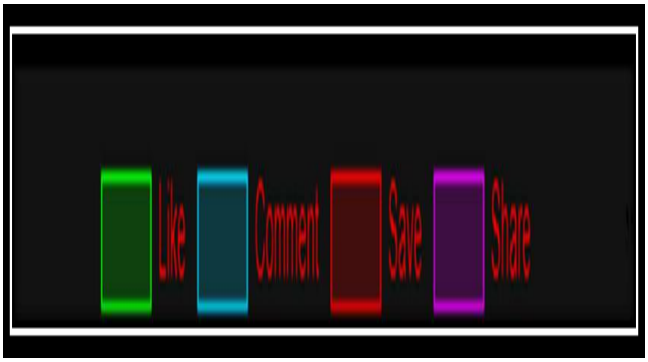

Figure 5. displays color illustrations of engagements.

Figure 6 breaks down the percentages of engagement for each individual post into the following categories: follower, shares, save, likes, and comments. The figure presents these categories. The document supplies a quantitative representation of each post's proportion and marks the posts with precisely established colors.

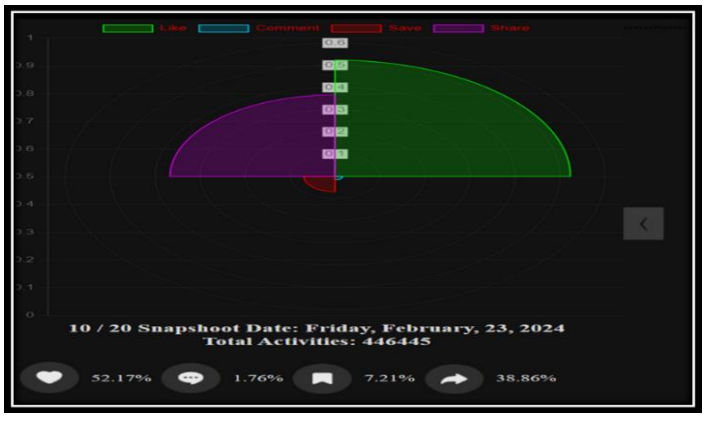

Figure 6. breaks down the percentages

In Figure 7 presents a visual representation of the aggregated and evaluated total number of posts, comments, saves, shares, and likes.

DOI: https://doi.org/10.54216/FPA.180203
Received: August 08, 2024 Revised: November 01, 2024 Accepted: January 11, 2025

Figure 7. presents a visual representation of the aggregated and evaluated

4. Conclusion

This study concludes with reflections and future directions, having analyzed the characteristics of 87,724 posts by prominent persons across various nations worldwide. The characteristics derived from the data by integrating post attributes and user features. This research aims to provide tips for improving social media popularity through the analysis of post features. Alongside the categorization of posts and topics, these attributes encompass the timeliness of publications, the correlation between features and metrics, and the classification of articles. It was established after thorough analysis that virality, engagement (play), and exposure (play/follower) were the three essential criteria. The investigation's findings suggest that postings on Friday and Saturday provide the potential visibility, but Sunday postings generate the maximum amount of engagement. Between noon and three o'clock, publication presents the optimal opportunity for complete recognition. Publications released before noon often garner greater engagement and have a higher likelihood of going viral, but those disseminated after noon tend to receive more attention than their earlier counterparts do. Users may choose to enhance either the exposure or the attraction metrics at their disposal prior to making a decision on the topic. The research results categories social media posts into five primary groups. Informal postings highlight inferior quality, focusing procedures rather than substance. Highlight volume posts about the number of contributions, whereas Supporting Posts focus on referencing or monitoring other users. Furthermore, contributions from influencers, characterized by a substantial following, are included as well. Informal postings generally achieve the greatest visibility. The standard characteristics of the postings consist of a caption averaging 828.2 characters, around 9.81 comments, and a tendency to favor commonly used hashtags rather than contextually relevant or personal hashtags. Correlation research and clustering indicate that shorter films, ideally 97.40 seconds or less, appear more desirable on some social media platforms. The discovery was made by the use of these two strategies. Users are anticipated to restrict the amount of accounts they follow and to uphold a succinct biography, preferably including of 1,091 followers and 960 characters. Excessive video sharing may enhance virality and profile visibility; but it will adversely affect exposure and attractiveness. Optimal exposure is achieved by producing roughly 161 posts daily and around 1,092 posts monthly.

References

[1] A. Alley and J. Hanshew, “A long article about short videos: A content analysis of US academic libraries’ use of TikTok,” *J. Acad. Librariansh.*, vol. 48, no. 6, p. 102611, 2022.

[2] K. Almgren, J. Lee, and M. Kim, “Prediction of image popularity over time on social media networks,” in *Proc. Ann. CT Conf. Ind. Electron., Technol. Autom. (CT-IETA)*, 2016, pp. 1–6.

[3] A. Fiallos, C. Fiallos, and S. Figueroa, “TikTok and education: Discovering knowledge through learning videos,” in *Proc. 8th Int. Conf. EDemocracy EGovernment (ICEDEG)*, 2021, pp. 172–176.

[4] A. R. Gómez, “Digital fame and fortune in the age of social media: A classification of social media influencers,” *aDResearch: Rev. Int. Invest. Comunicación*, no. 19, pp. 8–29, 2019.

[5] C. H. Basch, G. C. Hillyer, and C. Jaime, “COVID-19 on TikTok: Harnessing an emerging social media platform to convey important public health messages,” *Int. J. Adolesc. Med. Health*, vol. 34, no. 5, pp. 367–369, 2022.

[6] M. Bilal, N. Malik, N. Bashir, M. Marjani, I. A. T. Hashem, and A. Gani, “Profiling social media campaigns and political influence: The case of Pakistani politics,” in *Proc. 13th Int. Conf. Math., Actuarial Sci., Comput. Sci., Stat. (MACS)*, 2019, pp. 1–7.

DOI: https://doi.org/10.54216/FPA.180203
Received: August 08, 2024 Revised: November 01, 2024 Accepted: January 11, 2025

[7] H. Boholano, "Smart social networking: 21st century teaching and learning skills," *Res. Pedagogy*, vol. 7, no. 1, pp. 21–29, 2017.

[8] S. M. Chan-Olmsted, M. Cho, and S. Lee, "User perceptions of social media: A comparative study of perceived characteristics and user profiles by social media," *Online J. Commun. Media Technol.*, vol. 3, no. 4, pp. 149–178, 2013.

[9] A. Chopra, V. Avhad, and S. Jaju, "Influencer marketing: An exploratory study to identify antecedents of consumer behavior of millennials," *Bus. Perspect. Res.*, vol. 9, no. 1, pp. 77–91, 2021.

[10] A. W. P. Daryus, R. B. Ahmad, and M. Dada, "The factors influencing the popularity of TikTok among Generation Z: A quantitative study in Yogyakarta, Indonesia," *Electron. J. Bus. Manage* vol. 7, no. 1, pp. 1380–2550, 2022.

[11] C. Dhaoui and C. M. Webster, "Brand and consumer engagement behaviors on Facebook brand pages: Let's have a (positive) conversation," *Int. J. Res. Mark.*, vol. 38, no. 1, pp. 155–175, 2021.

[12] C. I. Eke, A. A. Norman, L. Shuib, and H. F. Nweke, "A survey of user profiling: State-of-the-art, challenges, and solutions," *IEEE Access*, vol. 7, pp. 144907–144924, 2019.

[13] O. P. John and S. Srivastava, "The Big-Five trait taxonomy: History, measurement, and theoretical perspectives," 1999.

[14] G. Kazai, I. Yusof, and D. Clarke, "Personalised news and blog recommendations based on user location, Facebook and Twitter user profiling," in *Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2016, pp. 1129–1132.

[15] L. Sun, H. Zhang, S. Zhang, and J. Luo, "Content-based analysis of the cultural differences between TikTok and Douyin," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, 2020, pp. 4779–4786.

[16] T. Sun et al., "Vaping on TikTok: A systematic thematic analysis," *Tobacco Control*, vol. 32, no. 2, pp. 251–254, 2023.

[17] D.-C. Tomozei and L. Massoulié, "Distributed user profiling via spectral methods," *Stochastic Syst.*, vol. 4, no. 1, pp. 1–43, 2014.

[18] M. Yadav, "Social media as a marketing tool: Opportunities and challenges," *Indian J. Mark.*, vol. 47, no. 3, pp. 16–28, 2017.

[19] M.-Y. Kim, S. Moon, and D. Iacobucci, "The influence of global brand distribution on brand popularity on social media," *J. Int. Mark.*, vol. 27, no. 4, pp. 22–38, 2019.

[20] Y. Ma and Y. Hu, "Business model innovation and experimentation in transforming economies: ByteDance and TikTok," *Manage. Org. Rev.*, vol. 17, no. 2, pp. 382–388, 2021.

[21] A. Masciantonio, D. Bourguignon, P. Bouchat, M. Balty, and B. Rimé, "Don't put all social network sites in one basket: Facebook, Instagram, Twitter, TikTok, and their relations with well-being during the COVID-19 pandemic," *PLOS One*, vol. 16, no. 3, p. e0248384, 2021.

[22] R. Portera, F. Bonacina, A. Corsini, E. S. Miele, and L. Ricciardi Celsi, "Energy profiling of end-users in service and industry sectors with use of Complex Network Analysis," *Proc. Complex Netw. Anal.*, vol. 312, 2021, pp. 1–16.

[23] K. R. Purba and Y. Yulia, "Realistic influence maximization based on followers score and engagement grade on Instagram," *Bull. Electr. Eng. Inform*, vol. 10, no. 2, pp. 1046–1053, 2021.

[24] Z. Qiyang and H. Jung, "Learning and sharing creative skills with short videos: A case study of user behavior in TikTok and Bilibili," in *Int. Assoc. Soc. Des. Res. Conf.*, 2019, vol. 10, pp. 25–50.

[25] S. Dixon, "Global social networks ranked by number," *Statista*, 2022. [Online]. Available: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

[26] A. Shutsko, "User-generated short video content in social media. A case study of TikTok," in *Proc. Soc. Comput. Social Media. Participation, User Exp., Consum. Exp. Soc. Comput.*, 2020, pp. 108–125.

[27] L. Yuan, H. Xia, and Q. Ye, "The effect of advertising strategies on a short video platform: Evidence from TikTok," *Ind. Manage. Data Syst.*, vol. 122, no. 8, pp. 1956–1974, 2022.

[28] Y.-Q. Zhu, D. Amelina, and D. C. Yen, "Celebrity endorsement and impulsive buying intentions in social commerce—The case of Instagram in Indonesia," in *Res. Anthol. Social Media Advertising Consum. Relat*, IGI Global, 2022, pp. 1401–1419.

DOI: https://doi.org/10.54216/FPA.180203
Received: August 08, 2024 Revised: November 01, 2024 Accepted: January 11, 2025