



DOI: <https://doi.org/10.38035/dijemss.v7i2>
<https://creativecommons.org/licenses/by/4.0/>

Discovering Hidden Pattern in Social Media Data to Empower MSMEs in The Digital Economy

Indira Nuansa Ratri¹, Arum Andary Ratri², Junaedi Adi Prasetyo³

¹Digital Business Study Program, Banyuwangi State Polytechnic, Indonesia, indira.nuansaratri@poliwangi.ac.id

²Digital Business Study Program, Banyuwangi State Polytechnic, Indonesia, arumandaryratri@poliwangi.ac.id

³Computer Engineering Technology Study Program, Banyuwangi State Polytechnic, Indonesia, junaedi.prasetyo@poliwangi.ac.id

Corresponding Author: indira.nuansaratri@poliwangi.ac.id¹

Abstract: The digital economy presents new opportunities and challenges for micro, small, and medium enterprises (MSMEs). This study explores hidden patterns in social media data to strengthen MSMEs in the culinary, tourism, and fashion-textile sectors. Using a mixed-methods approach, quantitative data mining of leading brands' social media was combined with qualitative validation through interviews with nine MSME representatives. Quantitative analysis revealed key engagement trends and content strategies, while qualitative insights confirmed their relevance and feasibility for SMEs. The integration of findings demonstrates that social media analytics can guide SMEs in enhancing digital presence, improving competitiveness, and adapting to shifting consumer behaviors. The study contributes practical recommendations for MSMEs while positioning social media as a strategic tool for sustainable growth in the digital era.

Keywords: Digital economy, MSMEs, social media analytics, hidden patterns.

INTRODUCTION

Micro, Small, and Medium Enterprises (MSMEs) are the backbone of Indonesia's economy, contributing more than 60% to the gross domestic product (GDP) and absorbing approximately 97% of the national workforce. Digital transformation has become the key for MSMEs to remain competitive in the digital economy era. Ministry of Cooperatives and SMEs (2024), by July 2024, around 25.5 million MSMEs had entered the digital ecosystem, reflecting a significant increase in technology adoption among small businesses. Furthermore, a survey by the Institute for Development of Economics and Finance (INDEF, 2023) noted that 56.3% of MSMEs actively conduct sales through social media platforms such as Instagram, Facebook, and TikTok, surpassing the utilization of e-commerce platforms, which stood at only 47.64%. These findings highlight the vital role of social media, particularly Instagram, in accelerating the growth of MSMEs within the digital landscape.

Instagram, as a visual and interaction-based platform, offers diverse features such as Instagram Shop, Reels, and Stories, which can be effectively utilized for marketing and

customer engagement. Its ability to deliver messages in both visual and narrative forms makes it a powerful promotional medium for building brand awareness and fostering consumer interaction. However, most MSMEs remain limited to the use of basic features and have not yet fully leveraged the potential of social media data analytics. Through social media analytics, businesses can better understand customer preferences, identify market trends, and evaluate content effectiveness, thereby enabling data-driven marketing strategies.

Several previous studies have emphasized the significance of social media adoption in MSME business strategies. Rahadi and Abdillah (2013) indicated that social media provides cost-effective promotion with broad reach and direct interaction with customers. Panjaitan et al. (2023) demonstrated that implementing digital marketing strategies through social media significantly improves sales performance. Efendioglu and Durmaz (2022) further found that consumer perceptions of Instagram advertising positively influence advertising value and brand awareness, which subsequently affect purchase decisions. In addition, Hammar et al. (2019) revealed that in-depth analysis of Instagram content can uncover hidden attributes relevant to enhancing product recommendation systems.

Despite these valuable contributions, prior research has predominantly focused on the general benefits of social media as promotional and communication tools, with limited exploration into the extraction of hidden patterns from user interactions. In particular, studies examining the complex interplay between posting time, content type, hashtags, sentiment in comments, and their influence on conversion outcomes within the Indonesian MSME context remain scarce. This gap underscores the need for more comprehensive investigations that move beyond descriptive use toward deeper analytical approaches capable of identifying latent behavioral patterns in social media data.

Aligned with the rapid growth of the digital economy supported by big data, internet penetration, and artificial intelligence, MSMEs have considerable opportunities to expand their markets and customize products according to consumer needs. Nevertheless, challenges such as limited digital literacy and inadequate data analysis capacity hinder MSMEs from fully optimizing these opportunities. A collaborative study by the Boston Consulting Group, Blibli, and Kompas Data (2024) projected that increasing MSME digital literacy from 20% to 50% could boost Indonesia's economy by up to USD 38 billion. Consequently, optimizing the use of social media data is not only important but also strategic for national competitiveness.

In response to these challenges, the present study aims to discover hidden patterns from Instagram user interaction data—including hashtags, comments, posting times, and content types—that can be applied by MSMEs in designing more effective digital content strategies. The novelty of this research lies in its integrative approach to social media analytics, which not only describes content performance but also extracts latent patterns to support MSMEs in enhancing customer engagement, expanding their reach, and improving digital sales conversion in the digital economy era.

METHOD

The research applies a sequential explanatory design, combining quantitative data analytics and qualitative interviews. Quantitative analysis is first conducted to extract hidden patterns from established Instagram brand accounts. The subsequent qualitative stage involves semi-structured interviews with SMEs to validate and contextualize the findings. In the first stage, quantitative analysis is carried out by extracting and analyzing Instagram data from established brand accounts to identify hidden patterns in engagement and content strategies. In the second stage, qualitative methods are employed through semi-structured interviews with SMEs to validate the applicability of these patterns in real-world contexts. This design ensures that quantitative findings are not only statistically valid but also relevant to the lived experiences of SMEs in the culinary, tourism, and fashion and textile sectors, which dominate the SME landscape in Indonesia. The following figure illustrates the methodology.



Figure 1. Sequential Explanatory Design

Data Source and Sampling

Quantitative Source

Instagram accounts of large and verified brands operating in culinary, tourism, and fashion–textile industries are purposively selected. The criteria for selection include:

Table 1. Indicators of Instagram Brand Account

No	Indicator	Description
1	National or International Brand	Brands with more than 100,000 followers
2	Engagement Rate (ER)	High engagement rate (measured as likes + comments divided by total followers) > 1%.
3	Consistent Content	Content that is consistently scheduled (at least 3 times per week) and the use of non-spam hashtags.
4	Visual Identity and Branding	Consistent Instagram accounts with coherent tone colors, design style, and caption style.
5	Verified Status	Verified account (blue check mark).

Qualitative Source

SMEs in culinary, tourism, and fashion and textile industries who actively use Instagram for marketing are invited to participate in semi-structured interviews. The purposive sampling ensures that the validation process directly reflects the challenges and opportunities faced by the largest SME sectors in Indonesia.

Data Collection

Instagram Data Scraping

Instagram data is extracted using Python-based scraping tools. Variables collected include engagement metrics (likes, comments, shares, saves), content type (photo, video, reels, carousel), caption length and sentiment, hashtag usage, posting time, and follower counts.

Semi-structured Interviews

Interview guidelines are prepared to explore themes listed in Table 2:

Table 2. Survey Questions for MSMEs (Micro, Small, and Medium Enterprises)

No	Aspect	Question
1	Instagram Usage Profile	Please explain how your MSME utilizes Instagram in marketing activities?
2	Type of Content	What types of content do you usually upload on Instagram? (Photos, videos, stories, etc.)
3	Posting Time	How do you determine the best time to upload content?
4	Perception of Posting Patterns	Based on our analysis, we found that video content with specific hashtags and posting times generates higher engagement. What is your opinion on this?
5	Experience with Features	Have you ever tried using specific hashtags and video content in your marketing strategy? If yes, what were the results? If not, what were the obstacles?

6	Implementation Challenges	In your opinion, what are the main challenges in applying digital marketing patterns similar to those used by larger brands?
7	Support and Training	What kind of support or information would you need the most to improve the effectiveness of your MSME's digital marketing?
8	Role of Technology and Data	How do you perceive the role of technology and data analytics in supporting the future development of micro, small, and medium enterprises?
9	Additional Feedback	Do you have any suggestions or additional input related to improving Instagram marketing performance for MSMEs like yours?

Data Analysis

Pre-processing

Data cleaning and normalization are applied to remove duplicates, handle missing values, and standardize variables such as posting times and text.

Hidden Pattern Discovery:

Clustering analysis is used to group content with similar engagement performance ; *Text mining and sentiment analysis* are applied to captions and comments to evaluate audience responses; Comparative analysis highlights optimal content formats, posting schedules, and hashtag strategies.

Validation with SMEs

Interview data is *analyzed* thematically to assess whether the discovered patterns are feasible and relevant for SMEs in culinary, tourism, and fashion and textile sectors.

Integration of Findings

Quantitative results are triangulated with qualitative insights. This integration ensures that discovered *patterns* are not only statistically significant but also practically actionable for SMEs. The ultimate goal is to formulate recommendations that enhance SMEs' competitiveness in the digital economy.

RESULTS AND DISCUSSION

Result

This section presents the findings obtained from both the quantitative and qualitative stages of the research. The results are structured to provide a clear understanding of hidden patterns in Instagram content and their implications for empowering SMEs in the digital economy. Quantitative findings are summarized through descriptive statistics and visualizations to highlight content preferences and posting behaviors. Meanwhile, qualitative insights from interviews with SMEs in the culinary, tourism, and fashion and textile sectors are used to validate and enrich the quantitative patterns. The first section presents the descriptive results (quantitative stage) of the social media analytics, focusing on instagram data from culinary businesses to highlight key patterns in content performance.

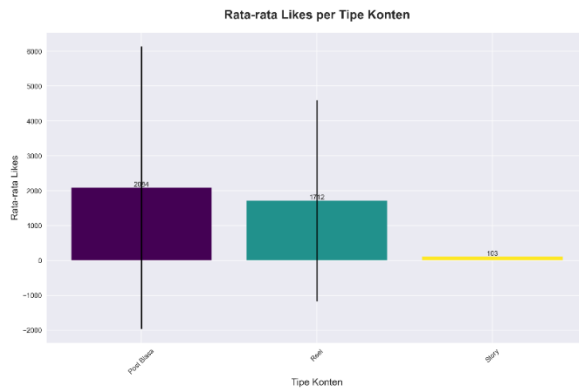


Figure 2. Mean Number of Likes by Content Type (Culinary)

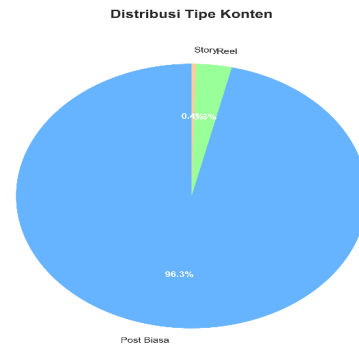


Figure 3. Content Type Distribution (Culinary)

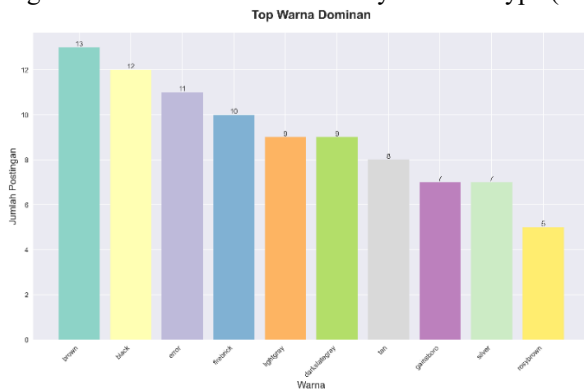


Figure 4. The Dominant Colors Utilized (Culinary)

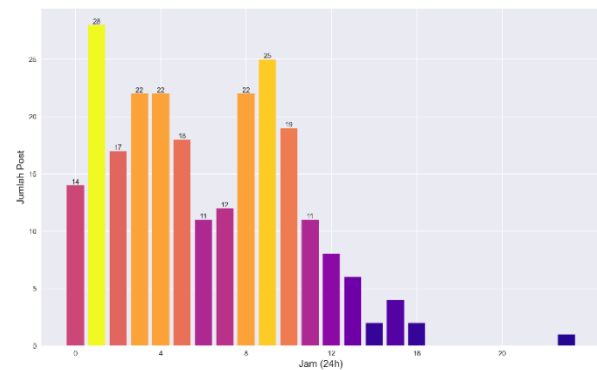


Figure 5. Distribution of Posting Hours (Culinary)

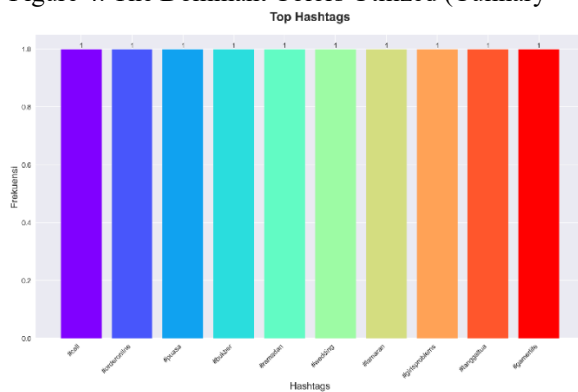


Figure 6. Most Frequently Used Hashtags (Culinary)

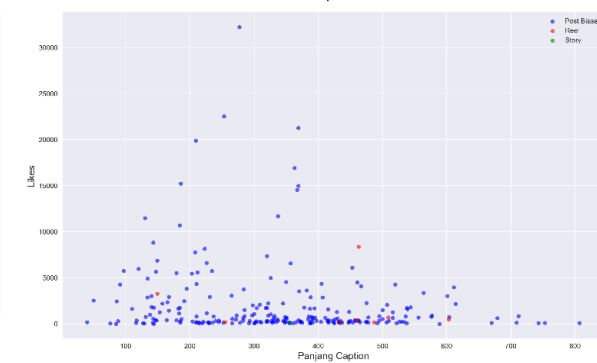


Figure 7. Most Frequently Used Hashtags (Culinary)

Figure 2 shows that static image posts (“regular posts”) achieved the highest average likes ($\approx 2,084$ likes/post), followed by Reels ($\approx 1,712$ likes/post), while Stories trailed behind (≈ 103 likes/post). Contrary to claims that short-form videos dominate, still images remain highly effective in the culinary niche. Figure 3 shows that the overall portfolio is heavily skewed toward static posts ($\approx 96.3\%$), with minimal Reels ($\approx 3.3\%$) and Stories ($\approx 0.4\%$). This suggests static images are not only effective but also preferred by brands. Figure 4 shows that color extraction identified brown, black, brick-red, light/dark grays, tan, silver, and rosy brown as the most frequent dominant hues. These align with food photography conventions, emphasizing texture, warmth, and clarity. Figure 5 shows that posting activity clustered around late night and morning hours, with peaks at 01:00 ($n=28$) and 09:00 ($n=25$). Engagement strategy thus appears to target early-day exposure, though time zone variation is possible. Figure 6 shows that Top hashtags were rotated contextually, including cultural (e.g., #ramadan, #bukber), event/lifecycle (e.g., #wedding, #lamaran), and functional tags (#orderonline). Figure 6 illustrates that large brands avoid repetitive, generic lists. Figure 7 shows that scatterplot analysis shows no strong linear correlation between caption length and likes. While some high-

Occasional late-night posts occurred at 23:00 (see Figure 11). Across the sampled tourism accounts, the top-10 hashtags showed low repetition: the two most frequent tags occurred 2 times each, whereas the remaining eight appeared once each (range = 1–2; median = 1). Thus, no hashtag appeared more than twice within the top-10 set (see Figure 12). Caption length varied broadly (≈ 200 –1,050 characters), while likes *ranged* from near zero to roughly 140,000. Most observations clustered below 5,000 likes across caption lengths. A small number of outliers with higher likes ($\geq 30,000$) appeared at mid-length captions (≈ 300 –450 characters). Visual inspection indicated no clear linear association between caption length and likes in this sample. Points labeled as Reels (red) were present among higher-like outliers; Stories (green) were sparse and near the lower-like region (see Figure 13). The next analysis focuses on the fashion–textile industries.

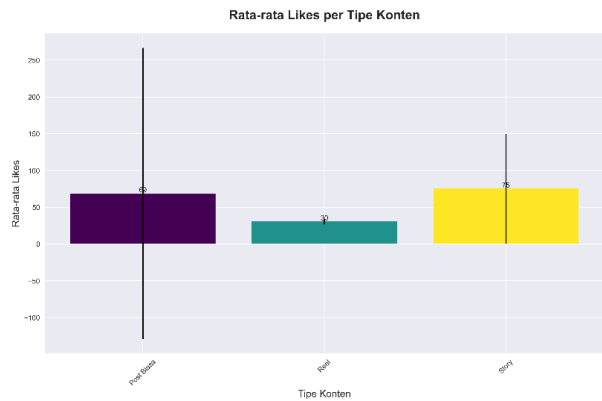


Figure 14. Mean Number of Likes by Content Type (Textile)

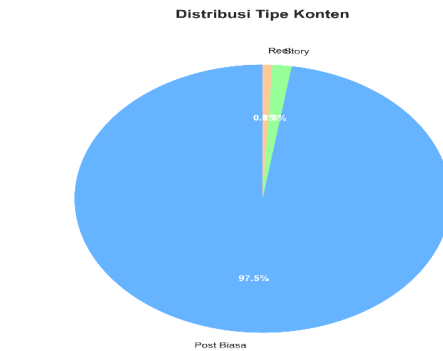


Figure 15. Content Type Distribution (Textile)

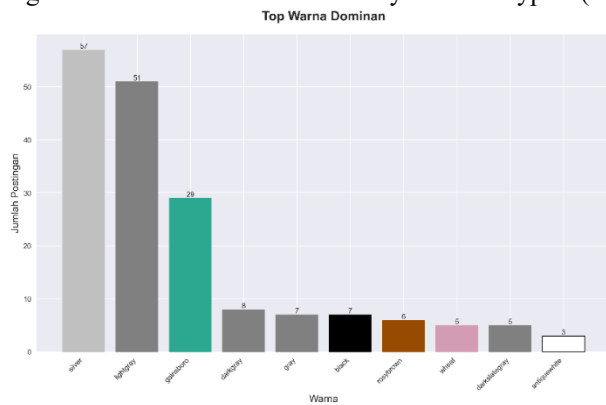


Figure 16. The Dominant Colors Utilized (Textile)

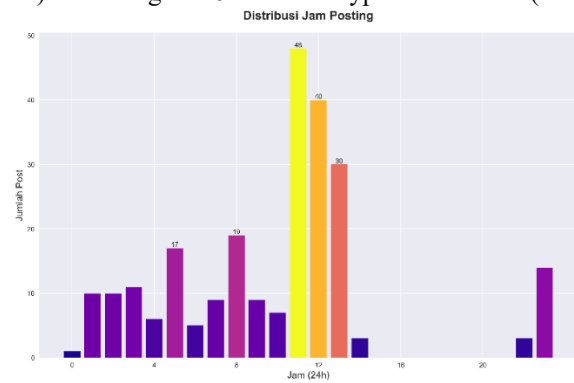


Figure 17. Distribution of Posting Hours (Textile)

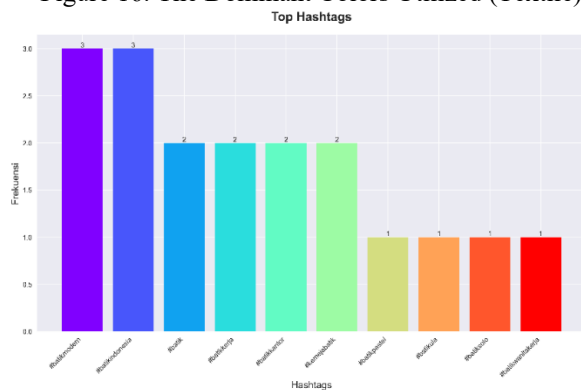


Figure 18. Most Frequently Used Hashtags (Textile)

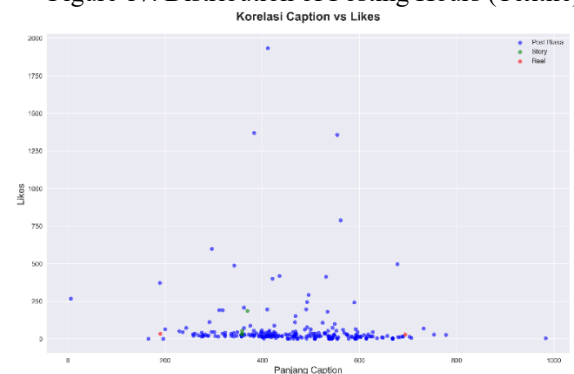


Figure 19. Most Frequently Used Hashtags (Textile)

Across the five benchmark accounts, Stories yielded the highest mean likes, followed by standard feed posts, while Reels showed the lowest averages (Figure 14). In absolute terms, Stories were only modestly ahead of posts, and Reels trailed clearly. These differences indicate a weak advantage for Stories over posts in this sample, with short-form video underperforming.

The content portfolio was overwhelmingly dominated by standard posts ($\approx 97.5\%$), with Stories ($\sim 1.7\%$) and Reels ($\sim 0.8\%$) appearing rarely (Figure 15). Thus, the observed performance patterns above are drawn from a feed-post-heavy strategy with limited short-form video deployment. The modal palette was neutral/achromatic, led by silver, lightgray, and gainsboro, each occurring far more often than any saturated hue (Figure 16). Smaller counts appeared for darker neutrals (e.g., darkgray, dimgray, black) and a few muted colors (e.g., rosybrown, wheat). Overall, the accounts favored a minimalist, desaturated aesthetic typical of premium fashion branding. Posting activity clustered in late morning to midday. The highest frequencies occurred at 10:00–12:00, with secondary early-morning activity ($\approx 04:00$ – $08:00$) and very limited posting after mid-afternoon (Figure 17). This pattern suggests a narrow “prime window” around late morning across the benchmark. Hashtags were concentrated on generic fashion discoverability. The most frequent tags were #ootd and #fashionkorea (each appearing three times), followed by #style, #batik, #streetwear, and several adjacent style/retail tags (each two occurrences; Figure 18). Niche, product-specific, or campaign-coded tags were comparatively rare. A weak, non-linear association was observed between caption length and likes (Figure 19). Likes remained low across most lengths, with a few higher-like outliers in the ~ 300 – 550 -character range for feed posts. Extremely short or very long captions did not systematically correspond to higher likes, and Reels/Stories were too sparse to infer format-specific trends.

To complement the quantitative findings, in-depth interviews were conducted with nine SME owners representing three sectors: culinary ($n = 3$), tourism ($n = 3$), and fashion and textile ($n = 3$). The purpose of these interviews was to validate whether the identified digital marketing patterns—particularly Instagram posting behavior, content type, and feature utilization—are applicable and relevant in real business contexts. Overall, participants confirmed the relevance of the quantitative results. The majority agreed that video content combined with appropriate hashtags and strategic posting times generates higher engagement, consistent with the statistical patterns. One participant from the tourism sector stated that “videos with clear storytelling and location tags attract more viewers compared to photos alone.” Similarly, fashion and textile SMEs highlighted that hashtag combinations help them reach niche markets, particularly younger audiences. Regarding posting time, respondents noted that evenings and weekends tend to attract more interactions, which aligns with the engagement peaks identified in the quantitative phase. However, culinary SMEs emphasized that timing may also depend on meal times, suggesting contextual nuances across sectors. In terms of implementation challenges, participants reported limited human resources, lack of advanced digital marketing skills, and difficulties in keeping up with platform algorithm changes. These insights provide depth to the statistical findings by revealing practical barriers faced in daily operations. When asked about support and training needs, SMEs expressed a strong demand for workshops on content creation, data-driven marketing, and analytics tools. This highlights that while the patterns are feasible, their effectiveness depends on SMEs’ access to capacity-building programs. In summary, the qualitative validation supports the statistical results and affirms that the discovered patterns are both feasible and actionable for SMEs across the three sectors, while also revealing contextual challenges and needs for further support.

Discussion

The findings of this study highlight the potential of uncovering hidden patterns in social media data as a strategic tool for empowering MSMEs in the digital economy. While quantitative analysis revealed statistically significant posting patterns and content types, the qualitative validation with MSMEs emphasized the practical feasibility of these insights. This integration underlines a critical point: data-driven strategies become meaningful only when they align with the lived realities of SMEs operating under resource constraints. The importance of this research lies in demonstrating how advanced data mining approaches, typically associated with large corporations, can be contextualized for small and medium enterprises. Prior studies

have shown that MSMEs often lack the analytical capacity to transform raw digital data into actionable insights (Nguyen et al., 2021; Troise & Tani, 2020). By simplifying these patterns into practical recommendations—such as optimal posting times, content formats, and the strategic use of hashtags—this study contributes to narrowing the digital divide between MSMEs and larger firms.

From a broader perspective, the empowerment of MSMEs through data-driven marketing is integral to fostering resilience in the digital economy. The digital economy thrives on information flows and consumer engagement (Bukht & Heeks, 2018). When MSMEs adopt evidence-based practices, they not only increase their competitiveness but also strengthen local economic ecosystems by participating more effectively in digital marketplaces. This is consistent with the Resource-Based View (RBV), which posits that firms can achieve sustainable advantage by leveraging unique, data-driven capabilities (Barney, 1991). Furthermore, the results demonstrate that while MSMEs in culinary, tourism, and fashion sectors differ in their operational contexts, they share common challenges in digital marketing adoption—particularly limited skills and support systems. This finding underscores the necessity of institutional interventions, such as training and digital literacy programs, to complement the insights generated by data mining. Without such support, the potential of social media analytics to empower MSMEs remains underutilized (Chatterjee & Kar, 2020).

In sum, this research contributes to the growing body of knowledge on digital entrepreneurship by illustrating how hidden patterns in social media data can be systematically discovered and translated into actionable strategies for MSMEs. More importantly, it demonstrates that empowerment in the digital economy is not merely about access to technology, but about transforming data into meaningful practices that align with the operational realities of small enterprises. This has both theoretical implications for understanding data-driven competitiveness and practical implications for policy and capacity-building programs aimed at MSMEs.

CONCLUSION

This study demonstrates that discovering hidden patterns in social media data can provide actionable insights to empower MSMEs in the digital economy, particularly in the culinary, tourism, and fashion sectors. By integrating quantitative analysis with qualitative validation, the research shows that data-driven strategies are both feasible and relevant for MSMEs, enabling them to optimize digital marketing practices and strengthen competitiveness. The findings suggest that the practical application of social media analytics requires not only technological tools but also capacity-building efforts to ensure MSMEs can translate insights into sustainable business practices. Future research and policy initiatives should therefore focus on developing accessible analytics frameworks, training programs, and institutional support systems that bridge the gap between data availability and effective utilization, ultimately fostering inclusive participation of MSMEs in the digital economy.

REFERENCE

- Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99–120.
- Bukht, R., & Heeks, R. (2018). Defining, conceptualising and measuring the digital economy. *International Organisations Research Journal*, 13(2), 143–172
- Chatterjee, S., & Kar, A. K. (2020). Why do small and medium enterprises use social media marketing and what is the impact: Empirical insights from India. *International Journal of Information Management*, 53, 102103.
- Efendioglu, A. M., & Durmaz, Y. (2022). The effect of Instagram advertisements on advertising value, brand awareness, and purchase intention. *International Journal of Business and Management Invention*, 11(2), 20–28.

- Hammar, E., Karlsen, R., & Syversen, T. (2019). Identifying hidden attributes in Instagram text for product recommendations. *Journal of Retailing and Consumer Services*, 51, 302–310.
- INDEF. (2023). *Digital economy survey report*. Institute for Development of Economics and Finance.
- Ministry of Cooperatives and SMEs. (2024). *Annual report on MSME digital transformation*. Jakarta: Kemenkop UKM.
- Nguyen, T. H., Newby, M., & Macaulay, M. J. (2021). SMEs' adoption of social media for marketing: A systematic literature review. *Journal of Small Business and Enterprise Development*, 28(2), 257–282.
- Panjaitan, N., Simanjuntak, M., & Pramudito, A. (2023). Social media-based digital marketing and its impact on SME performance. *Journal of Business and Entrepreneurship*, 15(1), 45–56.
- Rahadi, D. R., & Abdillah, L. A. (2013). Social media adoption in micro and small enterprises: Benefits and challenges. *Journal of Information Systems*, 9(1), 15–22.
- The Boston Consulting Group, Blibli, & Kompas Data. (2024). *Digital literacy and MSME economic growth in Indonesia*. Jakarta: BCG.
- Troise, C., & Tani, M. (2020). Exploring entrepreneurial characteristics, motivations and behaviours in equity crowdfunding: Some evidence from Italy. *Journal of Small Business and Enterprise Development*, 27(1), 183–203.