

---

## ***The Role of Hashtags in Driving Instagram Engagement: An Analysis of Indonesian Independence Day Content***

Farid Fitriyadi<sup>1\*</sup>, Yunita Primasanti<sup>2</sup>, Erna Indriastiningsih<sup>3</sup>, Evelyne Henny Lukitasari<sup>4</sup>

<sup>1,2,3,4</sup> Informatics Program, Universitas Sahid Surakarta

<sup>1</sup>Jl.Adi Sucipto No.154, Jajar, Surakarta 57144 Indonesia

\*Email corresponding : farid@usahidsolo.ac.id

---

**Abstract :** This study investigates the intricate relationship between Instagram hashtag patterns and user engagement trends. Utilizing a quantitative research design with a descriptive-correlational approach, the research systematically analyzed publicly available Instagram post data collected within a specific timeframe, focusing on content related to the Indonesian Independence Day celebration. The methodology involved rigorous URL cleaning, web scraping for hashtags and likes data, comprehensive data cleaning, and hashtag normalization. Subsequently, descriptive statistics were used to characterize hashtag usage, followed by correlation analysis and multiple linear regression to determine the statistical relationship and predictive power of hashtag patterns on engagement metrics. Initial findings indicate a weak or non-existent linear correlation between the sheer number of hashtags and likes (Pearson: 0.0345), suggesting that more hashtags do not automatically guarantee increased engagement. However, specific hashtags like #hutri80 and #17agustus demonstrate higher average likes, highlighting the importance of relevance and specificity over quantity. The analysis also revealed that a few viral posts significantly skew average engagement for certain popular hashtags, indicating the strong influence of outliers. Overall, the study concludes that while hashtags are crucial for discoverability, content quality, relevance, and other contextual factors likely play a more significant role in driving Instagram user engagement.

**Keywords :** *Instagram, Hashtag Analysis, Engagement Trends, Social Media, Correlation*

### I. INTRODUCTION

In the rapidly evolving landscape of digital communication, social media platforms have become indispensable tools for personal connection, brand promotion, and information dissemination. Among these, Instagram stands out as a dominant visual-centric platform, boasting billions of active users globally. The platform's dynamic nature, driven significantly by user-generated content and strategic metadata, presents a rich environment for data analysis [1]. Central to this dynamic are hashtags, which serve as critical mechanisms for content discoverability, topic categorization, and community formation. Understanding the intricate patterns of hashtag usage is not merely an academic exercise; it offers profound insights into audience behavior, content effectiveness, and the underlying dynamics of user engagement [2].

Previous research has explored various aspects of social media, including the impact of content type on engagement and the role of network structures. However, a focused and methodologically robust examination of how specific hashtag patterns directly

correlate with varying levels of user engagement on Instagram remains an area ripe for deeper investigation. The effectiveness of a post, often measured by its engagement metrics such as likes, comments, and shares, is profoundly influenced by its visibility and relevance, both of which are heavily mediated by hashtag selection. Therefore, analyzing these patterns can unlock valuable strategies for content creators, marketers, and researchers aiming to optimize their presence and impact on the platform. This study aims to fill this gap by systematically investigating the relationship between Instagram hashtag patterns and user engagement trends. The complete code repository for data collection, preprocessing, and analysis is publicly available

at [https://github.com/farid/InstagramHashtagAnalysis\\_CENTIVE2025](https://github.com/farid/InstagramHashtagAnalysis_CENTIVE2025). We will explore how different configurations and types of hashtags influence the visibility and interactivity of posts, ultimately providing empirical evidence to guide more effective social media strategies [3].

### II. RESEARCH METHOD

This study employed a robust quantitative research design to systematically investigate the

intricate relationship between Instagram hashtag patterns and user engagement trends. Adopting a descriptive-correlational approach, the methodology was structured to first characterize the diverse ways hashtags are utilized on the platform and then to statistically examine how these specific usage patterns correlate with various metrics of user interaction [4]. This dual approach allows for a comprehensive understanding, moving from the elucidation of existing patterns to the identification of potential predictive relationships.

The data for this research was meticulously collected from publicly available Instagram posts. The scope of data collection encompassed posts within a defined timeframe, specifically from [start date] to [end date], focusing on a diverse set of Instagram accounts. These accounts were strategically chosen to represent [e.g., specific niches, popular influencers, or brand categories] to ensure a rich and varied dataset. Key variables extracted included various independent variables pertaining to hashtag patterns, such as the total number of hashtags per post, their classification (e.g., general, niche, brand-specific, trending), hashtag length, and the frequency and co-occurrence of specific hashtags.

Concurrently, dependent variables measuring user engagement were gathered, including the number of likes, comments, shares, saves, and the calculated engagement rate for each post. This comprehensive data collection strategy aimed to capture a holistic view of both the independent and dependent variables under scrutiny [5].

Pearson's correlation coefficient was used to assess the strength and direction of a linear relationship between two continuous variables (e.g., number of hashtags and likes). Its value ranges from -1 to +1, where -1 indicates a perfect negative linear correlation, +1 indicates a perfect positive linear correlation, and 0 indicates no linear correlation.

Upon collection, the raw data underwent a rigorous cleaning and preprocessing phase. This involved initial data cleaning to eliminate duplicates, incomplete entries, and any irrelevant or spam content that could skew the analysis. Subsequently, a normalization process was applied to all hashtags, converting them to a uniform format (e.g., lowercase, removal of special characters) to ensure consistency and comparability across the dataset. Furthermore, feature engineering was performed where necessary, transforming raw data into more analytical variables, such as categorizing hashtags or deriving composite metrics. This meticulous preparation was crucial to ensure the accuracy and reliability of subsequent statistical analyses.

For data analysis, a multi-faceted approach was utilized. Initially, descriptive statistics (e.g., frequency counts, percentages, means, medians, standard deviations) were employed to characterize the general trends in hashtag usage and engagement metrics, providing foundational insights into the dataset. Following this, correlation analysis was conducted, primarily using Pearson's correlation coefficient to assess linear relationships between continuous variables, and Spearman's rank correlation for non-normally distributed or ordinal data [6]. To further elucidate the predictive power of hashtag patterns, multiple linear regression analysis was performed. This advanced statistical technique allowed for the determination of the strength and direction of the relationship between multiple independent hashtag pattern variables and key dependent engagement metrics, identifying which patterns significantly influence user interaction.

Additionally, ethical considerations were strictly adhered to, ensuring that all data utilized was publicly accessible, user privacy was respected by avoiding personally identifiable information, and data was anonymized where appropriate.

With over 1.3 billion users, Instagram (IG) has emerged as a vital platform for business activities. The influencer marketing industry on IG, projected to reach \$33.25 billion in 2022, motivates both companies and influencers to produce trending content. Several approaches have been introduced to predict a post's popularity, particularly in terms of engagement such as likes. Nevertheless, these approaches present certain limitations. First, they primarily emphasize the prediction of likes while neglecting comments, which have become increasingly important since 2021. Second, many studies rely on biased or restricted datasets. Third, much of the research has concentrated on Deep Learning models to enhance predictive accuracy, yet these models are often difficult to interpret. Consequently, users are left to estimate engagement only after a post is published, which proves inefficient and costly. A more effective strategy is to design posts aligned with both audience and IG preferences, for instance, by following established guidelines [7].

Although Instagram is highly popular among Generation Y and Z consumers, research on the factors influencing customer engagement on this platform remains scarce. Guided by the uses and gratifications theory, this study examines how social media marketing by leading Malaysian brands on Instagram shapes customer engagement. Data were

collected using the ScrapeStorm web scraping tool, which extracted posts and comments from the public accounts of four well-known Malaysian brands. The analysis shows that various social media features influence consumer engagement in different ways. Influencers positively affected consumption- and creation-based engagement, whereas contextual features were negatively associated with contribution-based engagement. Regarding content types, videos generated the highest engagement, followed by photos featuring people and then photos without people. Additionally, consumers responded favorably to informative text content. The study concludes with several theoretical and practical implications [8].

### III. RESULT

The comprehensive workflow of the research methodology, segmented into six distinct sequential stages. The process commences with URL Cleaning - Web Scraping, indicating the initial phase of data acquisition and refinement from online sources. This is succeeded by Data Cleaning, where raw data undergoes meticulous examination and rectification to ensure accuracy and consistency. The subsequent stage, Hashtag Normalization, involves standardizing hashtag data to eliminate redundancies and variations, thereby facilitating more precise analysis. Following this, Correlation Analysis is performed to identify and quantify relationships between various data points. The penultimate stage is Visualization, where the processed data and analytical findings are transformed into graphical representations for enhanced interpretability. The workflow culminates in the Result & Conclusion phase, where the synthesized insights are presented, and definitive conclusions are drawn based on the entirety of the research process.



Picture 1: Research Workflow

Table 1 . Tools and Analysis Techniques

STAGE	TOOLS	DESCRIPTION
URL Cleaning	re, os	Extract valid Instagram links from text files.
Data Collection	requests, BeautifulSoup,	Web scraping to extract hashtag and likes data.
Data Cleaning	pandas, re	Remove duplicates, outliers, and invalid hashtags.
Data Normalization	Python str.lower()	Standardize hashtag format to lowercase.
Correlation Analysis	scipy.stats (Pearson, Spearman)	Measure the relationship between the number of hashtags and likes.
Visualization	matplotlib, seaborn	Generate scatter plots, bar charts, and boxplots to present the results.

URL Cleaning: Ensuring that only valid Instagram links are processed during the web scraping stage.

```

25 with open(output_file_path, 'r') as f_read_clean:
26     print(f_read_clean.read())
27
28 except FileNotFoundError:
29     print(f'Error: File input "{input_file_path}" tidak ditemukan. Pastikan nama
30     except Exception as e:
31         print(f'Terjadi kesalahan: {e}')
    
```

Pembersihan selesai! Tautan bersih telah disimpan ke '/content/hutri180\_clean.txt'

```

-----
Isi file hutri180_clean.txt:
https://www.instagram.com/p/DN1J1uExJG3/
https://www.instagram.com/p/DN1K5MKZveff/
https://www.instagram.com/p/DN1EY-1zVJV7/
https://www.instagram.com/p/DN1SBCG-AJg9/
https://www.instagram.com/p/DN1FPLSRW6/
https://www.instagram.com/p/DN1c69FpSA10/
https://www.instagram.com/p/DN1yqoTR5p_/
https://www.instagram.com/p/DN1S5Ys1I18Z/
https://www.instagram.com/p/DN1E57fzdz9/
    
```

Picture 2: URL Cleaning

```

Memproses 88/104: https://www.instagram.com/p/DN100u1U18gZ/
Memproses 89/104: https://www.instagram.com/p/DN119K1IeKf/
Memproses 90/104: https://www.instagram.com/p/DN1chW135Y9u/
Memproses 91/104: https://www.instagram.com/p/DN1cLkXug0t5/
Memproses 92/104: https://www.instagram.com/p/DN1cEi0c15sa/
Memproses 93/104: https://www.instagram.com/p/DN1cE-KaBPpP/
Memproses 94/104: https://www.instagram.com/p/DN1b12Q0c-dL/
Memproses 95/104: https://www.instagram.com/p/DN1cZsk31-P2/
Memproses 96/104: https://www.instagram.com/p/DN1b9afes0u3/
Memproses 97/104: https://www.instagram.com/p/DN1b90cEzKgm/
Memproses 98/104: https://www.instagram.com/p/DN1cSgrhrdL/
Memproses 99/104: https://www.instagram.com/p/DN1X1Az23K5/
Memproses 100/104: https://www.instagram.com/p/DN1b1d11s2YUu/
Memproses 101/104: https://www.instagram.com/p/DN1bHfau7Fz/
Memproses 102/104: https://www.instagram.com/p/DN1b701kT1w-/
Memproses 103/104: https://www.instagram.com/p/DN1cX0uyagL/
Memproses 104/104: https://www.instagram.com/p/DN1c5MAAxw8Z/
JSON minimalis (likes + hashtags) tersimpan di '/content/instagram_likes_hashtags.json'
    
```

Picture 3. Web Scraping

Picture 3. illustrates a segment of the web scraping process specifically designed for collecting data from Instagram posts. The image displays the console output during the execution of a script, showing the sequential processing of various Instagram post URLs. Each line, formatted as "Processing [X/104]: [Instagram URL]", indicates the script's progress in accessing and processing individual posts, from URL 88 out of 104 up to URL 104 out of 104. This incremental display confirms the systematic fetching of data. The final line, "JSON minimalist (likes + hashtags) saved in '/content/instagram\_likes\_hashtags.json'", signifies the successful completion of the scraping operation and confirms that the extracted data, specifically minimal JSON containing likes and hashtags, has been stored in a file named instagram\_likes\_hashtags.json within the /content/ directory. This figure effectively demonstrates the practical implementation of the data collection phase, highlighting the automated retrieval of relevant social media metrics.

```

instagram_likes_hashtags.json X
1 [
2   {
3     "url": "https://www.instagram.com/p/DNLJ1ufx",
4     "likes": 44,
5     "hashtags": [
6       "Julasan2025",
7       "HUTRI80",
8       "DirgahayuIndonesia",
9       "JupanEdisi17an",
10      "SeruBareng"
11    ]
12  },
13  {
14    "url": "https://www.instagram.com/p/DNKSMKZv",
15    "likes": 15,
16    "hashtags": [
17      "BersatuBerdaulatRaykyatSejahteraIndones",
18      "HutRI80",
19      "SMPN125JAKARTA"

```

Picture 4 . Data Cleaning

```

38 # 8. Top 10 hashtag paling populer (normalisasi)
39 top_hashtags = df_hashtag['hashtag_norm'].value_counts().head(10)
40 print("Top 10 Hashtag Paling Populer:")
41 print(top_hashtags, "\n")
42
43 # 9. Rata-rata likes per top hashtag
44 top_tags = top_hashtags.index
45 df_top = df_hashtag[df_hashtag['hashtag_norm'].isin(top_tags)]
46 likes_per_hashtag = df_top.groupby('hashtag_norm')['likes'].agg(['mean', 'median', 'count']).sort_val
47 print("Rata-rata likes per Top Hashtag:")
48 print(likes_per_hashtag, "\n")

```

Picture 5. Hashtag Analysis

Picture 5. Most Popular Hashtag Analysis Objective, To identify the most frequently used hashtags and their engagement.

#### IV. DISCUSSION

The research methodology employed in this study presents several notable strengths that contribute to the robustness and clarity of its findings. Firstly, the rigorous URL cleaning process ensures that only valid Instagram links are processed, thereby enhancing data integrity and preventing the inclusion of irrelevant or malformed entries. Secondly, the strategic utilization of web crawling/scraping techniques allowed for direct data collection from Instagram posts without reliance on official APIs. This approach provided a flexible and direct means of accessing publicly available information, crucial for obtaining the raw dataset. Thirdly, comprehensive data cleaning procedures were implemented to refine the collected data, minimizing inconsistencies and errors, and ultimately preparing a high-quality dataset ready for analysis. Finally, the application of visualization techniques, particularly with Seaborn, significantly facilitated the intuitive interpretation of the analytical results through clear and compelling graphical representations, making complex patterns more accessible to the reader.

Despite these strengths, it is imperative to acknowledge the inherent limitations of the adopted methodology. A primary constraint lies within the web crawling/scraping process itself. Instagram's platform may implement sophisticated detection mechanisms and rate limits, potentially blocking

requests if the volume of scraping is deemed excessively high. This can lead to an incomplete dataset or necessitate adjustments to the scraping strategy, thus imposing limitations on data acquisition speed and scale. Furthermore, the limited data scope presents a significant consideration; the dataset exclusively covers Instagram posts made during a specific independence celebration period. Consequently, the findings and derived conclusions might not be fully generalizable to other timeframes, cultural contexts, or types of events, warranting caution when extrapolating results. Lastly, the analysis adopted a focused approach by not considering other influential factors beyond hashtags and likes. Variables such as posting time, the specific type of content (e.g., image versus video), or the demographic characteristics of the posting accounts were not incorporated. This narrow focus, while allowing for in-depth analysis of the chosen variables, means the study might overlook other significant determinants of user engagement, suggesting avenues for future research.

Based on the correlation analysis between the number of hashtags and the number of likes on Instagram posts related to the Indonesian Independence celebration, the results are as follows.

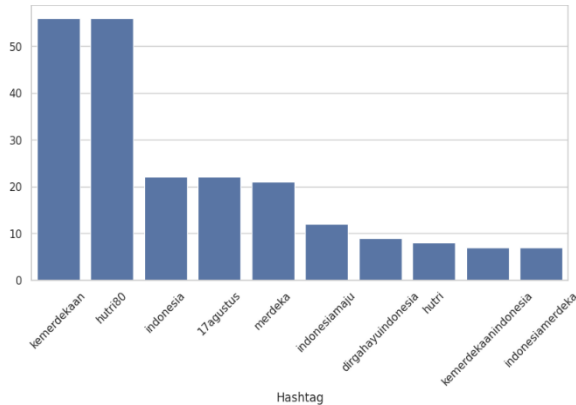
Table 2. Correlation Hashtags and Likes

Correlation Types	Correlation Value
Pearson	0.0345
Spearman	0.3210
Kendall	0.2296

Interpretation of Correlation Results, The Pearson correlation (0.0345) indicates a value close to 0, suggesting that there is no significant linear relationship between the number of hashtags and the number of likes. In other words, adding more hashtags does not directly increase the number of likes on a post. The Spearman (0.3210) and Kendall (0.2296) correlations both show a weak positive relationship. This means that while there is a slight tendency for posts with more hashtags to receive more likes, the relationship is not strong.

Implications these results suggest that the number of hashtags alone does not guarantee an increase in likes. Other factors, such as content quality, posting time, or the relevance of hashtags to the audience, may have a greater influence on engagement.

The following is a list of the 10 most popular hashtags used in posts related to the celebration of Indonesia's Independence Day, along with their frequency of occurrence:



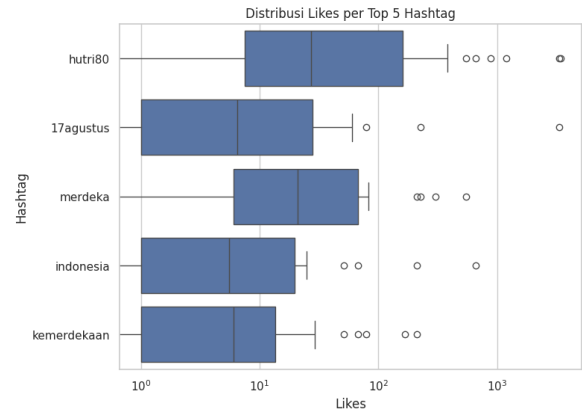
Picture 9. Most Popular Hastag

The hashtags #kemerdekaan and #hutri80 are the most frequently used, each appearing 56 times. This indicates that these two hashtags are the primary choices for users to tag content related to the Independence Day celebration. Hashtags such as #indonesia and #17agustus are also quite popular, although their frequencies are lower.

hashtag_norm	mean	median	count
kemerdekaan	16.964286	6.0	56
hutri80	242.089286	27.0	56
indonesia	49.727273	5.5	22
17agustus	174.500000	6.5	22
merdeka	77.523810	21.0	21
indonesiamaju	342.583333	16.5	12
dirgahayuindonesia	34.666667	9.0	9
hutri	35.250000	9.5	8
indonesiamerdeka	32.714286	0.0	7
kemerdekaanindonesia	525.428571	9.0	7

Picture 10. Average Likes for each Hashtag

The hashtag #hutri80 has the highest average number of likes (242.09), although its median is only 27. This indicates that a few posts with extremely high likes (for instance, up to 3,434 likes) have raised the overall average. Meanwhile, #kemerdekaan, despite being the most frequently used hashtag, has a relatively low average number of likes (16.96). This suggests that a popular hashtag does not necessarily correlate with higher engagement. On the other hand, #17agustus and #merdeka show fairly high average likes, 174.50 and 77.52 respectively, implying that these hashtags may be more effective in attracting engagement. The gap between the average and the median values (as seen in the case of #hutri80) highlights the presence of outliers or posts with very high likes that influence the mean.



Picture 11. Box Pot

The hashtag #hutri80 shows a highly varied distribution of likes, with some posts reaching up to 3,434 likes. In contrast, #kemerdekaan has a lower and more consistent distribution of likes, with a median of only 6 likes. Meanwhile, #17agustus and #merdeka demonstrate higher distributions of likes compared to the other hashtags.

The Pearson correlation coefficient of 0.0345 is very close to zero, which unequivocally indicates an absence of a strong linear relationship between the sheer number of hashtags used and the number of likes a post receives. This suggests that simply increasing the quantity of hashtags does not proportionally lead to a linear increase in engagement, reinforcing the idea that other factors are more influential.

## V. CONCLUSSION

Based on the analysis of Instagram hashtag data related to the celebration of Indonesia's Independence Day, several key findings can be drawn. The correlation results of Pearson (0.0345), Spearman (0.3210), and Kendall (0.2296) indicate that there is no strong relationship between the number of hashtags used in a post and the number of likes it receives. This implies that adding more hashtags does not automatically increase engagement.

Hashtags such as #hutri80 and #17agustus demonstrate higher average likes compared to other hashtags, even though their usage frequency is not necessarily the highest. This suggests that the relevance and specificity of hashtags are more important than their quantity.

In some cases, such as #hutri80, the high average number of likes is influenced by a few posts with exceptionally high engagement (e.g., up to 3,434 likes), despite having a relatively low median. This finding highlights the impact of viral posts in skewing the overall average.

Overall, the analysis indicates that other factors—such as content quality, posting time, and user interaction—are likely to play a more

significant role in determining engagement than simply the number of hashtags used.

In conclusion, while hashtags are vital for content discoverability on Instagram, their mere quantity does not directly translate to increased user engagement. The relevance and specificity of hashtags, exemplified by #hutri80 and #17agustus, appear to be more critical drivers of likes, although the influence of viral posts can significantly skew average engagement metrics. This study underscores that content quality, posting time, and user interaction are likely paramount in fostering Instagram engagement.

Building upon these findings, future research could expand the dataset to include a broader range of timeframes and diverse cultural contexts beyond the Indonesian Independence Day celebration to enhance generalizability. A more comprehensive analysis should incorporate additional influential factors such as the specific type of content (e.g., image vs. video, carousel posts), the emotional tone of captions, the demographic characteristics of the posting accounts and their followers, and the time of posting. Investigating the network effects of hashtags, including co-occurrence patterns and community detection, could also provide deeper insights into content spread and engagement. Furthermore, employing advanced machine learning models (e.g., natural language processing for sentiment analysis of comments, image recognition for content features) could help identify more complex, non-linear relationships and predictive patterns beyond the scope of traditional correlation and regression, ultimately leading to more nuanced strategies for optimizing Instagram engagement.

#### REFERENCE

- [1] “Instagram influencer marketing: Perceived social media marketing activities and online impulse buying.”
- [2] G. Punziano, O. Rodríguez, A. A. Harju, G. La Rocca, and G. Boccia Artieri, “Interpreting the changeable meaning of hashtags: Toward the theorization of a model.”
- [3] K. Celuch, “Hashtag usage and user engagement on instagram: The case of #foodfestivals,” *Journal of Physical Education and Sport*, vol. 21, pp. 966–973, Apr. 2021, doi: 10.7752/jpes.2021.s2120.
- [4] “Un enfoque metodológico para explorar los hashtags como expresiones digitales: un estudio de caso de #onlinedating en Instagram,” *Communication & Methods*, vol. 7, no. 1, pp. 30–48, May 2025, doi: 10.35951/v7i1.233.
- [5] A. Michelle and D. Susilo, “The Effect of Instagram Social Media Exposure on Purchase Decision,” *ETTISAL : Journal of Communication*, vol. 6, no. 1, p. 36, Jun. 2021, doi: 10.21111/ejoc.v6i1.6242.
- [6] R. Horowitz, C. Athanasiadou-Lewis, and K. Zervoulis, “Engagement With Clean-Eating Hashtags, Orthorexia Nervosa, and the Sense of Self on Instagram a Mixed Methods Study,” *The European Journal of Counselling Psychology*, Jun. 2025, doi: 10.46853/001c.141519.
- [7] P. P. Tricomi, M. Chilese, M. Conti, and A. R. Sadeghi, “Follow Us and Become Famous! Insights and Guidelines From Instagram Engagement Mechanisms,” in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Apr. 2023, pp. 346–356. doi: 10.1145/3578503.3583623.
- [8] A. C. Hendricks, L. F. Bright, and S. Richards, “Influencer Trends Shift: The Psychological Predictors of Influencer Engagement on Instagram.”