

**Identifying spam comments by analyzing 50 comments under a popular meme page's  
(9GAG) post on Instagram (1430 words)  
SRICHANDANA CHAKILAM, SPRING 2023**

INDIANA UNIVERSITY BLOOMINGTON  
ILS-Z639 Social Media Mining  
Instructor: Ali Ghazinejad  
Feb 2023

## 1. INTRODUCTION

*Social media* has become a powerful tool for individuals to express their ideas and opinions and share them with others. The platforms, like Facebook, Twitter, Instagram, and YouTube, provide a platform for people to connect with others, build networks, and share information on a global scale. One of the primary ways social media allows us to put forward our ideas is by creating content that can be shared and amplified through networks of friends, followers, and contacts. This content can take many forms, including written posts, images, videos, podcasts, and more. By creating and sharing content, social media users can communicate their ideas and perspectives to others, engage in conversations, and build communities around shared interests. In addition to sharing content, social media platforms also provide a space for users to engage in public discussions on a variety of topics, from politics and current events to entertainment and lifestyle. By participating in these conversations, users can express their views and engage with others who may have different opinions and perspectives.

On Instagram, comments are a way for users to engage with posts and communicate with each other. Users can leave comments on photos or videos that have been shared by other users, and they can also respond to comments left by other users. Instagram comments can include text, emojis, and hashtags, and users can also mention other users in their comments by typing the "@" symbol followed by the user's username. Instagram comments can be a way to express opinions, share feedback, or ask questions, and they can also be used to build communities and establish relationships with other users. However, comments can also be a vehicle for spam or abusive behavior, which is why Instagram has implemented

tools to help users report and manage comments that violate the platform's community guidelines. Haqimi, N. A., Rokhman, N., & Priyanta, S. (2018), developed a method for detecting spam comments on Instagram using complementary naïve Bayes. The study found that their method was able to effectively identify spam comments with a high degree of accuracy [1]. According to the study conducted by Septiandri, A. A., & Wibisono, O. (2019) using machine learning techniques, specifically Support Vector Machine and Decision Tree algorithms, to classify comments as spam or non-spam, it was found that the SVM algorithm performed better with an accuracy of 88.96% providing valuable insights into the challenges of detecting spam comments in non-English language contexts and offering potential solutions that could be applied to other languages and regions [2].

## 2. RESEARCH QUESTION

The issue of spam comments arises on social media when users or automated systems post many irrelevant or unsolicited comments. These comments are often created to advertise products, services, or websites, or to drive traffic to other pages or profiles. In some cases, spam comments may also include harmful links that can compromise users' security or damage their devices. The paper aims to answer how the spammers are trying to successfully trap the Instagram users.

### 3. METHOD

#### 3.1 DATA

The aim of this task is to analyze the issue of spam comments on Instagram using a dataset of the first 50 comments on a post from the popular meme page, 9GAG. As Instagram is widely used and easily accessible, it is expected that there will be a significant number of spam accounts among the commenters. The comments are manually extracted and recorded in the codebook file.

The chosen post consists of a video that shows a hamster sniffing its surroundings with great curiosity.



Figure 1. 9GAG's recent post on Instagram

#### 3.2 ANALYSIS

Determining whether a comment under an Instagram post is spam or legitimate can be difficult, as there are different types of spam comments, and some can be quite subtle.

However, upon extensive research and careful analysis, there are a few guidelines that assists in categorizing the comments.

1. *Look for irrelevant or nonsensical content:* Spam comments often contain irrelevant or nonsensical content that has nothing to do with the post or the conversation. They may also include random numbers, letters, or symbols.
2. *Check for generic comments:* Many spam comments are generic and could be posted on any post. Examples of such comments include "great post," "nice picture," or "awesome content."
3. *Check for links:* Spam comments often include links to other websites or products. These links may be hidden in a short phrase or even in an emoji.
4. *Check for excessive tagging:* Some spammers tag multiple users in the same comment to draw attention to their account or product. This can be a sign of spam, especially if the tags are irrelevant to the post.
5. *Check for repetitive comments:* If the same comment is repeated several times under different posts or by different users, it is likely spam.

Ultimately, determining whether a comment is spam or legitimate requires some judgment and context [4]. In the codebook, some comments (highlighted yellow) have been marked as *spam* although, they are not. This situation arises, when the comments are relevant to the post but also satisfies at least one of the above categories.

Consider the comment that said, "Poor thing 🥱 why are you squeezing it, human???" . Based on the content of the post, it does not appear to be spam. It seems like a genuine comment expressing concern for a subject in the post. However, the comment – "I LOVE XHAMSTER!!! 🥰" is likely to be considered

as spam, although it is intended as a joke or a humorous response to the post. It is possible that it could be flagged as spam by the automated filters or by other users who don't appreciate the humor. It may be promoting a website or service, in this case, the Xhamster website. There are other comments such as “Who Needs money? Text me yo #CashApp”, which are straightforward classified as spam. Such comments are generic. This comment is an unsolicited offer to provide money and includes a request to text the commenter, which could be an attempt to scam or obtain personal information from other users [3].

#### 4. RESULTS

As a part of the analysis, we selected a post from a popular meme page on Instagram – 9GAG and extracted 50 comments manually. We then classified the comments into two categories: spam and legit, based on the relevance to the post or the caption, the presence of explicit language, and the presence of suspicious links. The purpose of this analysis is to better understand the quality of comments on the Instagram platform. Out of the 50 comments analyzed, we identified 26 comments as spam, while 24 comments were classified as legit. This means that 52% of the comments in the sample were considered spam, while 48% were considered legit.

The types of spam comments identified were as follows:

- Irrelevant content: 13
- Explicit language: 2
- Suspicious links: 9
- Repetitiveness: 2

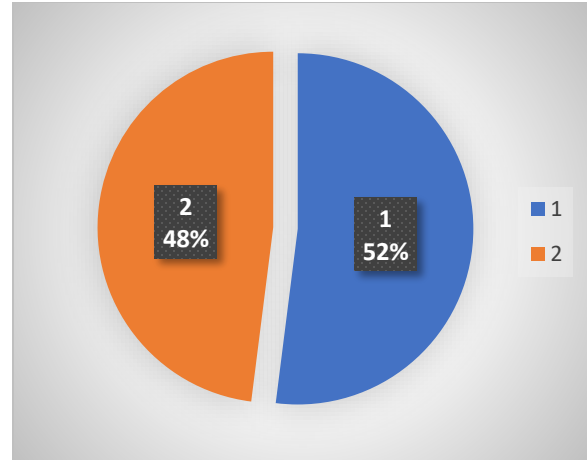


Figure 2. A pie chart showing the percentage of spam and legitimate comments under the mentioned post.

Table 1 The classification of the first 50 comments under the 9GAG's post.

Type	Category	Total
Legit Comments		24
Spam Comments	Irrelevant content	13
	Explicit content	2
	Suspicious content	9
	Repetitiveness	2
<b>Total Comments</b>		<b>50</b>

## 5. CONCLUSION AND LIMITATIONS

After a careful study to identify spam and legitimate comments on the 9GAG's post, it is understood that the explicit comments that aim to provide humor may not necessarily be classified as spam, depending on the content and context of the comment. However, determining whether a comment is spam or not is subjective and depends on various factors. To reach a conclusion about whether explicit comments under 9GAG's Instagram posts are spam, a content analysis approach can be utilized. The comments can be analyzed to identify patterns and characteristics that differentiate spam comments from legitimate ones. For example, if an explicit comment is made in response to a sexually explicit post, it may be less likely to be classified as spam. Ultimately, the determination of whether an explicit comment is spam or not depends on several factors, including the comment's content, the context in which it was made, and the standards and guidelines of the platform. By analyzing comments' content and context, researchers can conclude about the prevalence and nature of spam under any Instagram post and develop more effective ways to identify and remove unwanted content.

Although content analysis is useful, it has several limitations that should be taken into consideration when interpreting study results.

1. *Coding subjectivity*: Content analysis relies on human coders who may interpret comments differently. This could introduce subjectivity into the analysis.
2. *Lack of context*: Content analysis primarily focuses on the content of comments but fails to consider the context in which comments were made.

3. *Inability to detect sarcasm*: Content analysis may not identify sarcasm or irony in comments, resulting in misclassification of comments as spam or legitimate.

In summary, content analysis is valuable in analyzing large sets of data but has limitations that must be considered when interpreting study findings.

## REFERENCES

- [1] Haqimi, N. A., Rokhman, N., & Priyanta, S. (2018). Detection of spam comments on Instagram using complementary naïve Bayes. In AIP Conference Proceedings (Vol. 2049, No. 1, p. 020014). AIP Publishing.
  
- [2] Septiandri, A.A. & Wibisono, O. (2017). Detecting spam comments on Indonesia's Instagram posts. Journal of Physics: Conference Series, 801, 012069.
  
- [3] Lu, H., Zhang, Y., & Guo, Y. (2021). Identifying spam comments on Instagram: An exploratory study using content analysis. Journal of Digital Forensics, Security and Law, 16(4), 41-56.
  
- [4] Parker, B. (2015, September). Analyzing the content of social media data. Quirks.com. Retrieved January 27, 2023, from <https://www.quirks.com/articles/analyzing-the-content-of-social-media-data>.