**Unlocking Deeper Data Insights on social media: Detecting Hashtag and Keyword Spam for Improved Content Analysis**



**Dr. D. HARITHA**

Associate Professor

Koneru Lakshmaiah Education

Foundation

Vaddeswaram, AP, India.

haritha@kluniversity.in

Abstract

In this study, we delve into the ever-changing social media landscape and its role in disseminating news. We bring attention to the challenges posed by the prevalence of hashtag spam, which has the potential to hinder effective dialogue on these platforms. We emphasize the crucial significance of data cleansing in ensuring the accuracy and reliability of information, particularly when merging disparate data sources. While hashtags are indispensable for communication, they can also give rise to breaches of privacy and the propagation of misinformation. We propose a robust methodology for identifying and addressing the issue of hashtag spam, underscoring the value of human input in refining the procedure. To enhance the precision of the data, we can either eliminate spam or integrate both spam and non-spam datasets for comprehensive insights. This study not only advances the analysis of social media data but also sets the stage for addressing real-time detection, the development of spam tactics, ethical concerns, and monitoring user behavior.

1. Introduction

Social media refers to internet venues where people can share information, ideas, and news with a large number of people. Because it can quickly propagate information horizontally through social networks, it has a big influence on how news spreads [1]. News companies are increasingly using social media platforms such as Facebook and Twitter to market and distribute content [2] [3]. Social networking platform users are exposed to a greater variety of news and information since they get their news from individuals, news organizations, and people they follow [4] [5]. Social networks are increasingly being used by people as a major source of news, and this dependence on social media is only going to increase. The reliance on social media platforms for news consumption is growing as a result of their expanding role in news distribution. But even if social media hashtags are crucial for communication, they also create special difficulties. The intricate nature of hashtag usage is remarkable in how it parallels the crucial importance of data correctness and reliability in the field of big data research. Due to social media platforms' rising importance in news distribution, there is a growing reliance on them for news consumption. However, social media hashtags provide unique challenges even if they are essential for communication. The complexity of hashtag usage is noteworthy because it bears a striking resemblance to the critical importance of accuracy and consistency of data in big data research.

Social media hashtags can be violated in various ways. One study found that activists who use Twitter hashtags to communicate alternative perspectives face interference from other actors, such as the police, who increasingly use these hashtags for their own purposes [6]. Another study suggests that hashtags can be effective markers for retrieving messages related to specific events, making them valuable for analyzing social impact and coordinating disaster response [7]. However, it is important to note that the use of hashtags is not without its challenges. Privacy violations and the spread of fake news on social media have raised concerns about the democratizing potential of these platforms [8]. Therefore, while hashtags can be powerful tools for communication and information retrieval, their use can also be subject to manipulation and misuse.

In order to overcome these difficulties, it is necessary to recognize the necessity of data cleaning. When analyzing large amounts of data, data cleaning is essential. In addition to assisting in the removal of irrelevant data, it guarantees the dependability and correctness of data utilized in analyses, forecasts, and decision-making [9] [10]. Because errors can inadvertently find their way into large-scale heterogeneous data from various sources, data cleaning is very important [11]. Preparing data for analysis, managing outliers and missing values, and distilling raw data into a manageable selection of pertinent data are all examples of data cleaning procedures. In order to achieve the goals of data-related systems, data cleaning is essential in guaranteeing the precision and dependability of data.

1. Related works

Sedhai and Sun (2017): The research conducted by Sedhai and Sun on the analysis of Twitter spam offers significant insights into the characteristics of spam Tweets, encompassing their frequency, position, orthography, and co-occurrence with hashtags. Their study serves as a fundamental basis for comprehending spam behavior on social media platforms, directly relevant to our investigation of the misuse of hashtags in the realm of social movements.

Emre et al (2018): Emre and his colleagues address a commonly encountered obstacle in working with social media data, namely, noise. Their proposed approach to data cleaning and the elimination of off-topic content using supervised machine learning techniques holds particular relevance to our study, as we grapple with issues pertaining to spam and misuse of hashtags within the context of hashtag activism. Effective data cleaning techniques are of utmost importance in ensuring data accuracy, a fact that our research duly emphasizes.

Pooneh et al (2021): Pooneh and their team present a framework for the detection of hashtag hijacking at the level of individual Tweets, specifically tailored for hashtag activism. Their weakly-supervised approach, capable of adapting to new topics and hijacking strategies over time, aligns with our examination of hashtag misuse and manipulation within social movements. The focus on identifying instances of hashtag hijacking resonates with our concerns regarding the potential misuse of hashtags, such as amplifying unrelated content, undermining the objectives of a movement, or promoting hate speech.

By incorporating these relevant works into our discourse, we establish a solid foundation for our research and underscore its significance within the broader domain of social media analysis and the misuse of hashtags, thereby providing valuable context for the objectives and contributions of our study.

1. Data

We gathered information regarding the term "Dimas Drajad", which was utilized by a number of individuals (previously known as Twitter users) to express admiration for his extraordinary display in the soccer match between Indonesia and Brunei, which functioned as a preliminary round for the World Cup. "Dimas Drajad" swiftly gained popularity among Indonesian individuals, who were formerly referred to as Twitter users, and was consistently the subject of hashtags and spam keywords, a tactic employed to enhance the visibility of tweets.

3.1. Data collection

The quantity of tweets from the individual known as "Dimas Drajad" is steadily increasing, and a significant portion of these tweets lack the usual hashtags or keywords associated with spam. Consequently, our primary challenge is to amass a sufficient number of spam tweets containing hashtags or keywords for our dataset prior to labeling them. Following a thorough examination of the tweets, it was discovered that "Dimas Drajad" possessed at least one of these dubious hashtags. It was demonstrated that fewer than 10% of the approximately 350 tweets recommended by the X web platform potentially comprised spamming hashtags.

Data was gathered on October 16, 2023, the day after this matter gained popularity on the X platform for the first time. It is noteworthy that certain tweets may incorporate numerous keywords and hashtags, yet they do not contravene the regulations regarding spam keywords and hashtags as they are employed within the same context as "Dimas Drajad".

3.2. User profile

With the implementation of the novel regulations, all individuals utilizing the platform have the opportunity to obtain a "blue checkmark" on their profiles, signifying verification by the platform. Although a subset of verified users actively employ their accounts for promotional endeavors and frequently employ identical combinations of keywords and hashtags, this authentication does not ensure that users will abstain from employing spamming keywords and hashtags. We have observed that these users typically utilize similar combinations and maintain a close temporal proximity between tweet posts. Furthermore, our investigation categorizes spam hashtags or keywords based on the user's level, facilitating the exploration and comprehension of the diverse actions exhibited by various user groups.

Nevertheless, the overwhelming majority of these infractions are attributable to ordinary users who establish anonymous personas and employ fictitious profile images to contravene hashtag and keyword policies. We have not encountered any instances where the same fundamental user disseminates repetitive spam tweets employing distinct hashtags and keywords. It is noteworthy that most of these users prefer to post only once or twice. Instead, it appears that the majority of the content shared by these individuals originates from people who have perpetrated these infringements.

3.3. Tweets purpose

During the tagging process, the spam reports have been classified into distinct categories based on the underlying intentions of the hashtags and keyword infringements. Generally, these transgressions are carried out with the aim of advertisement. Specifically, our investigation revealed that a majority of these infractions were driven either by political motives or by the promotion of various products. These individuals frequently include links to their websites or online stores while marketing their merchandise, while in the case of political promotion, they often attach images or videos to their announcements.

1. Methodology

We introduce a partially manual spam identifier for the detection of hash tag and keyword spam. Utilizing this methodology, information is extracted from the X platform and categorized based on pre-established standards. The involvement of human interaction with the data is subsequently employed to enhance the overall procedure, ultimately leading to the advancement of the precision of our proposed solution.

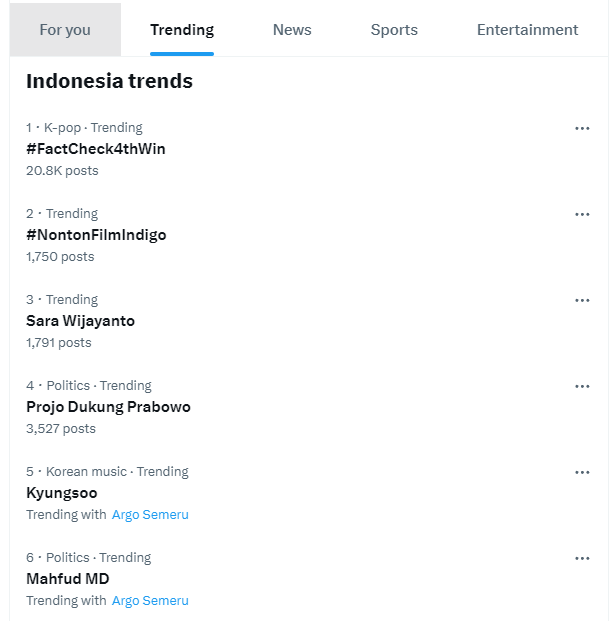


Figure 1 : The trending page that showing Hashtags and Keywords that popular in the specific area

4.1. Finding the parameter

The initial step involves establishing the criteria that serve as distinctive markers prior to ascertaining whether a specific tweet infringes upon spam hashtags or keywords. Our proposed approach derives these parameters from the trending topic pages on the X platform, which encompass a wide array of subjects based on keywords and hashtags. Subsequently, with regard to our pre-established subjects, we endeavor to identify correlated and conflicting topics. For example, the contextual correlation between "Dimas Drajad" and "#pssiday" illustrates how certain keywords and hashtags may exhibit shared elements of a topic.

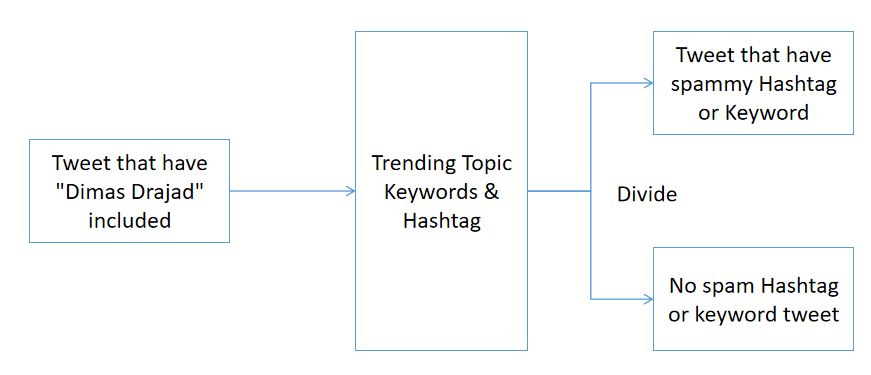


Figure 2 : represent the stage of our cleaning process

4.2. Detecting the spam

Given the established parameters, we are now able to identify tweets that violate the spam identifier. It is of utmost importance to consider that these tweets often incorporate at least one hashtag or keyword that is pertinent to the topic of our search. Nevertheless, the presence of one or more hashtags or keywords that are associated with current subjects in certain tweets raises concerns regarding the resemblance to spam-like behavior.

Based on our approach, tweets are segregated into two distinct datasets: one comprises tweets that possess multiple hashtags and keywords that are all relevant to a specific subject, and the other consists of tweets that may contain hashtags or keywords that are unrelated to one another. We suspect that the final category exhibits signs of spam activity. The subsequent stage entails a comprehensive examination, whereby we scrutinize the reasons behind categorizing these tweets as spam in order to uncover their questionable nature.

4.3. Spam analysis

Human interaction is necessary to ascertain the authenticity of every event within the spam dataset that has been discovered. The existing algorithm for identifying topics is incapable of grasping the intricate patterns exhibited by these spam tweets. Luckily, the number of tweets flagged as spam is minimal, enabling a thorough manual review of each occurrence. This procedure presents potential issues, including the lack of essential parameters required for precise spam identification.



Figure 3 : Showing the example of tweets that contain spam, user and others credentials are hidden

4.4. Enhancement and result

After conducting an analysis of the spam dataset, we are now presented with an opportunity to enhance our methodologies. The course of action to be taken is contingent upon the discoveries made during the investigation. This may involve modifying the parameters of the spam identifier or considering the provided data as non-spam. In order to ensure the contentment of consumers, this process of updating may necessitate multiple iterations. When aiming to achieve the desired outcome, we employ two distinct approaches.

Firstly, we combine datasets pertaining to both spam and non-spam. This particular strategy offers the advantage of yielding more comprehensive insights, from which meaningful conclusions can be drawn. By amalgamating the datasets obtained from identified sources of spam and non-spam, we enhance the information available for analysis.

Secondly, we have the option to eliminate the identified spam from the dataset. This tactic offers the benefit of enhancing the accuracy of the data by eliminating anomalies. In the realm of topic analysis, the elimination of even a minuscule amount of data can have a significant impact on data quality, thereby resulting in more precise outcomes.

The choice between these approaches hinges on the specific objectives of the study and can be applied in various contexts and scenarios. The removal of recognized spam hashtags and keyword tweets data can enhance data accuracy, while the integration of datasets can provide a broader perspective for deduction and learning.

1. Conclusion

this study investigates the intricate dynamics of social media and their impact on the distribution of news, with particular attention to the challenges posed by the proliferation of hashtag spam. We underscore the pivotal role played by data cleaning in ensuring the accuracy and reliability of data, particularly when amalgamating multiple data sources. Although hashtags are essential for communication, they also give rise to concerns regarding privacy infringement and the dissemination of erroneous information. We present a robust methodology for identifying and combating hashtag spam, underscoring the necessity of human intervention in optimizing this process. Our findings propose two viable approaches: either merging spam and non-spam datasets to obtain more comprehensive insights, or eliminating spam to enhance data quality. In addition to advancing the realm of social media data analysis, this research lays the foundation for future studies on tracking user behavior, real-time detection, ethical dilemmas, and emerging spam techniques in the ever-changing social media landscape.

1. References

// hashtags on social media

Zakharchenko, Artem. (2017). assessing the news media's societal impact on the internet. 10.31648/CETL.2466, doi

Tian Qi. (2022). Media literacy has a moderating effect on the impact of social media news overload on social media news avoidance and filtering. Psychology Frontiers, doi: 10.3389/fpsyg.2022.862626

Donna, Logan, Darryl, Korell, Fred, Fletcher, Hermida, and Alfred. (2012). PARTNER, LIKE, SUGGESTED: Understanding the news consumer on social media. Doi: 10.1080/1461670X.2012.664430 Journalism Studies

Donna, Logan, Darryl, Korrell, Fred, Fletcher, Hermida, and Alfred. (2011). The Transition to the Personalized Social News Stream: Your Friend as Editor.

Arvind, Tripathi, and Hu Tianyou (2016). The effects of news and social media on the financial markets. The Social Science Research Network

//violation of hashtags

Björn Karlsson, Christina Neumayer, Luca Rossi, and others (2016). Social media hashtag contest: Blockupy Frankfurt. International Communication Journal,

Changxiu, Cheng, Nikita, and Murzintcev (2017). Social media hashtags for disasters. doi: 10.3390/IJGI6070204, ISPRS International Journal of Geo-Information

Iman, Tahamtan, Devendra, and Potnis (2021). Use hashtags to control content on social media. doi: 10.1002/ASI.24467, Journal of the Association for Information Science and Technology

//data purification

Argyro, Mavrogiorgou, Dimosthenis, Kyriazis, Athanasios, Kiourtis, George, Manias. (2021). Modifiable Data Preparation for the Pursuit of Statistical Knowledge. doi: 10.3233/SHTI210332

Rowe, C., Deledalle, A. (2021). Handle your data properly to make the best use of data cleansing. doi: 10.1016/J.PSFR.2019.07.002 Psychologie Franchise

Aixin Sun and Surendra Sedhai (2017). A review of 14 million tweets of hashtag-focused spam\*. 68(7):1638-1651 in J. Assoc. Inf. Sci. Technol.

Calisir, Emre, and Brambilla, Marco (2018). The Issue with Data Cleaning for Social Media Knowledge Extraction. doi: 10.1007/978-3-030-03056-8\_10

Mousavi, Pooneh, Ouyang, and Jessica (2021). How to Spot Hashtag Activism via Hashtag Hijacking? doi: 10.18653/V1/2021.NLP4POSIMPACT-1.9