SAMI2021

Leveraging Phone Numbers for Spam detection in Online Social Networks

Authors: Rohit Jere, Anant Pandey, Manvi Singh, Mandar Ganjapurkar

Article ID: 28

Abstract

- Online Social Networks (OSNs) provides invaluable information to millions of individuals daily but has also become one of the most popular places for spam campaigns.

- Twitter, Flickr and Tumblr are a few examples of OSNs that provide a forum for the world to connect.

- While there are advantages of having these OSNs, they are still a long way to go in terms of spam detection and banning.

- In this paper, we design an algorithm for the recognition of spam campaigns, specifically focusing on a phone-numbers based approach.

Introduction

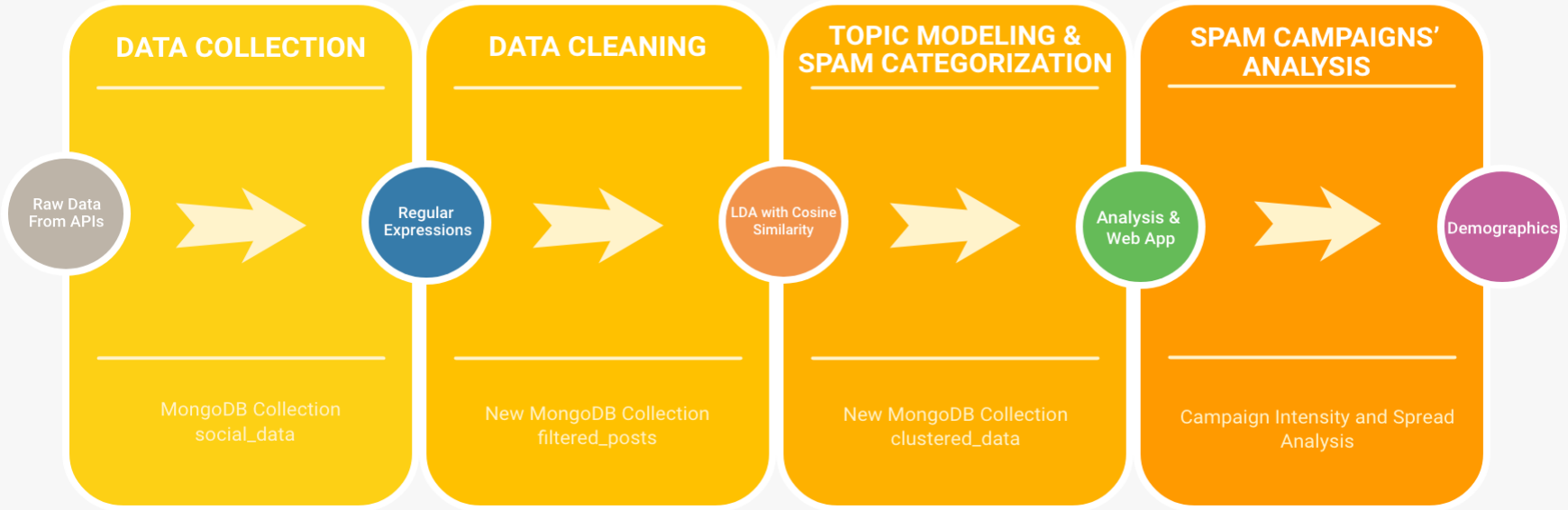
- Our paper presents a new study and analyzes the depth to which these spammers have corrupted today's popular OSNs. Our work extends through the following OSNs: Twitter, Flickr, and Tumblr. A large database with over 18 million posts was collected from these platforms collectively. The data took over 4 months to collect with the limited API support.

- These API scripts were run every few hours for a time frame of 4 months to collect over 18,000,000 posts from these platforms collectively.

- To collect data that potentially has a high hit ratio in terms of the spam on the internet, we collect this data using 400 keywords closely related to spam.

- A few of these keywords are as follows: call me, call us, 100%, #1, message, email, email me, email us, 100% free.

Methodology



Data Collection

Id: Unique ID of the post in question

RT(Retweet): This is a Boolean value which depends on whether a tweet is retweeted or not.

Keyword: Refers to the keyword used to perform the searching from the APIs.

Source: This property notes which OSN acts as the source of the post in question.

Content: This property notes the text that belongs to the potential spam post.


```
{
  "_id": {
    "$numberLong": "623227697704927232"
  },
  "keyword": "call me",
  "username": "alicedovey",
  "user_id": -1,
  "content": "Call me, on the line. Call me, call me, any
    , anytime. Call me... Oh, my love!When you're ready
    we can share the wine. Call me....",
  "timestamp": "10 Jul 2020",
  "location": -1,
  "hashtags": [
    "Blondie",
    "Call Me",
    "Music",
    "Songs",
    "Musica",
    "Song",
    "Video",
    "YouTube",
    "Pop",
    "Rock",
    "Love Songs"
  ],
  "Retweet": -1,
  "Fav": -1,
  "source": "tumblr",
  "date_added": "2020-07-10",
  "in_reply_to": -1
}
```

Data Collection

User_id: Unique ID of the user who posted the post in question

Timestamp: Refers to the point in time when the post was posted by the user.

In_reply_to: This property stores the unique ID of the user that the particular post in question was a reply to (if any). If not, this is hard-coded to -1.

Fav: This property stores the number of likes/dislikes the post in question received.

Date_added: This property refers to the timestamp. This post was collected and stored in our database.


```
{
  "_id": {
    "$numberLong": "623227697704927232"
  },
  "keyword": "call me",
  "username": "alicedovey",
  "user_id": -1,
  "content": "Call me, on the line. Call me, call me, any
    , anytime. Call me... Oh, my love!When you're ready
    we can share the wine. Call me....",
  "timestamp": "10 Jul 2020",
  "location": -1,
  "hashtags": [
    "Blondie",
    "Call Me",
    "Music",
    "Songs",
    "Musica",
    "Song",
    "Video",
    "YouTube",
    "Pop",
    "Rock",
    "Love Songs"
  ],
  "Retweet": -1,
  "Fav": -1,
  "source": "tumblr",
  "date_added": "2020-07-10",
  "in_reply_to": -1
}
```

Data Pruning

- Here, r1 represents the regular expression wherein we are looking for a number with 1 to 5 digits in the beginning which is superseded by another hyphen, and then superseded by another 3 to 7 digits.
- Here, notice that these phone numbers may appear both with or without country codes. Hence, we account for the same by creating another variable named r1p, wherein the 'p' represents that we are checking for country code alongside.
- Similarly, we take into account other popular phone number formats and check for numbers both with and without country codes. All of these in combination helps us in identifying 7 to 15-digit phone numbers as stipulated by ITU-T E. 164.

$$r1p = r'(\backslash + \backslash d\{1,5\} - \backslash d\{3,5\} - \backslash d\{3,7\})' \quad (1)$$

$$r2p = r'\backslash + \backslash(?[\backslash d]\{3\}\backslash)?[-\backslash s]?[\backslash d]\{3\}[-\backslash s]?[\backslash d]\{3,6\}' \quad (2)$$

$$r3p = r'(\backslash + \backslash(\backslash d\{3,5\}\backslash)[-\backslash s\backslash d]\backslash d\{2,5\}[-\backslash s\backslash d]\backslash d\{2,6\})' \quad (3)$$

$$rp = r'(\backslash + \backslash d[\backslash d\backslash s-]\{5,13\}\backslash d\{2,4\})' \quad (4)$$

Topic Modelling & Spam Categorization

- We started by pre-processing the pruned data. The following steps were taken: removal of stop words, stemming, and lemmatization. A global dictionary was created using this data, which was further changed into a corpus by leveraging gensim.

- A hybrid methodology using Latent Dirichlet Allocation (LDA) and Bag of Words (BoW) was articulated to classify this data into categories. This helped us obtain the top categories in our collected data.

- Probabilistic graphical model (PGM) formalization based LDA is a Bayesian model which can be implemented into other Bayesian models .

- It is based on PGM and can be used for gathering discrete data from domains including collaborative filtering and content-based image retrieval.

Topic Modelling & Spam Categorization

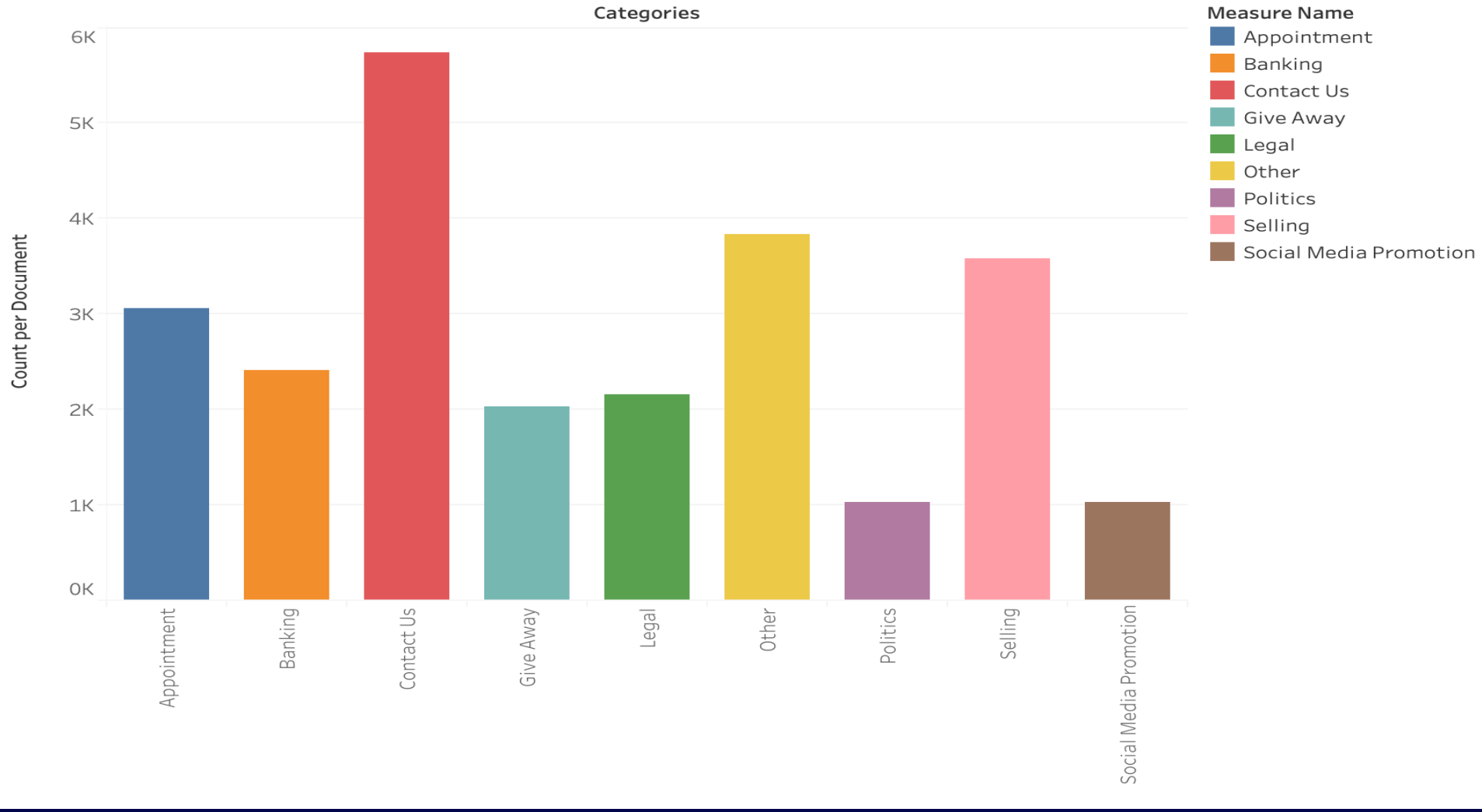
- Each of these latent topics generates words based on probability distribution. LDA is an example of topic modeling that is used for identifying abstract topics in documents. It categorizes texts into specific topics.

- Bag-of-words is used to abstract features from the text that involves measurement of the occurrence of known words present in a document. Precisely, it helps in determining the frequency of known words existing a document without paying heed to grammar and word order.

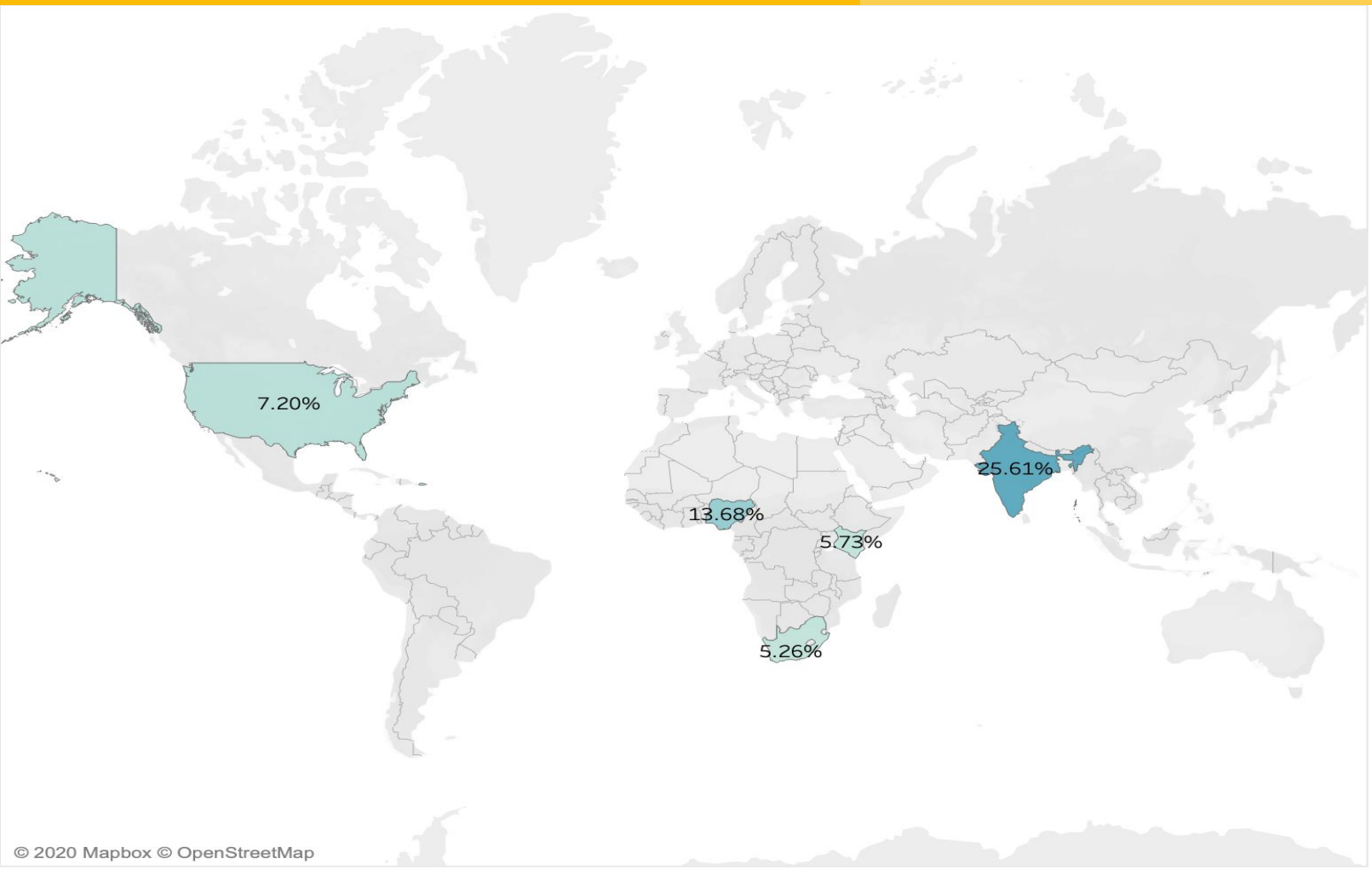
- Each post was sent into the class with the maximum similarity score. The following formula was used:

$$\cos(t, e) = \frac{te}{\|t\|e\|} = \frac{\sum_{i=1}^{n} t_i e_i}{\sqrt{\sum_{i=1}^{n} (t_i)^2} \sqrt{\sum_{i=1}^{n} (e_i)^2}}$$

Results and Analysis

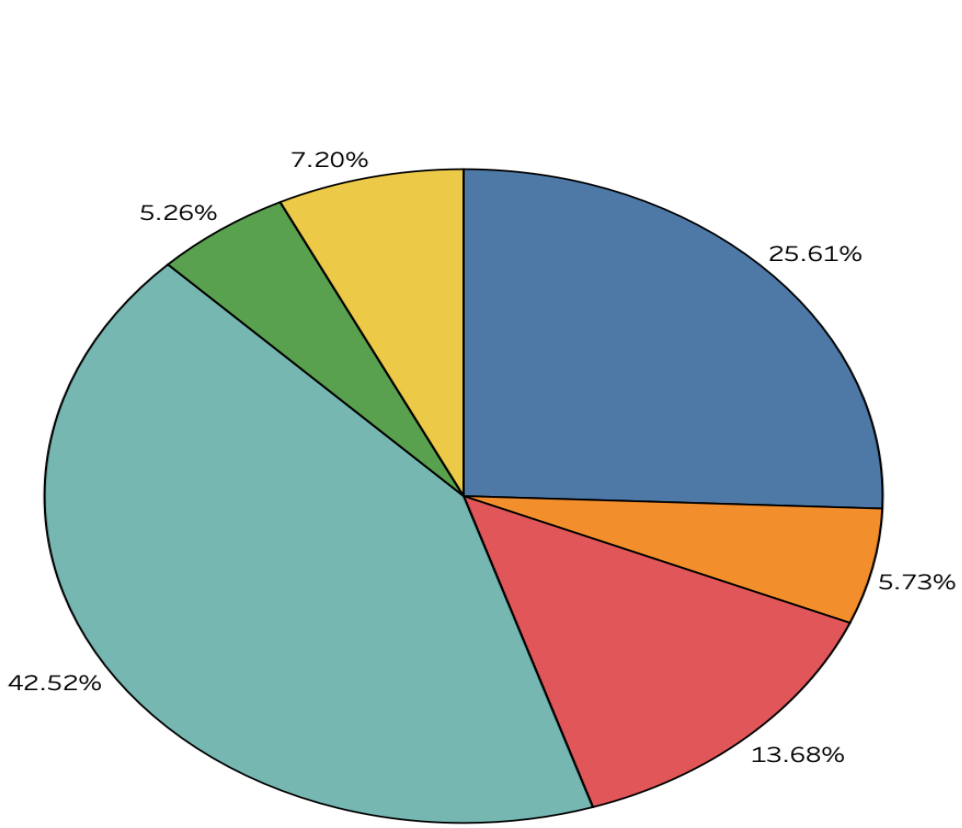


Results and Analysis



Map based on Longitude (generated) and Latitude (generated). Color shows % of Total Count. Details are shown for Country.

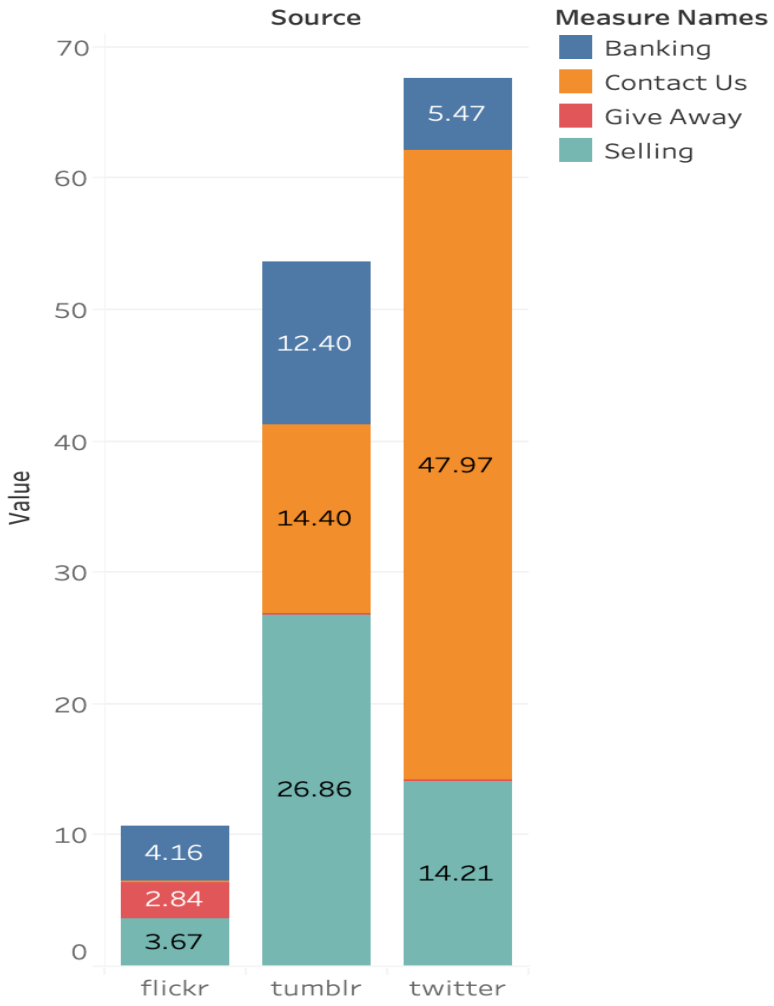
Results and Analysis



Results and Analysis

Category	Count
sweepstakes	4,074
instantwin	1,372
COVID19	1,311
URGENT	1,307
jobs	1,091
business	1,087
NiteFlirt	1,080
video	1,035
realestate	1,015
SpreadTheWord	994
RESCUE	944

Count: 944 – 4,074



Conclusion

- Social media, while being a huge asset in today's world, has also attracted its fair share of spam online. Spam fetters the ability of these platforms and potentially drives the rightful users into traps and negative experiences. For these platforms to thrive, spam on these OSNs needs to be controlled and fully minimized.

- Hence, the aim remains to identify and ban these spam posts while ensuring that no legitimate posts are banned. In this manner, the OSNs will be a safer place to transact information. We will fortify the trust of the common man in OSNs and help us build a safer, secure, and trustable environment where these spammers do not scam legitimate users. We will be able to identify such spam posts in this research and analyze the regions with the maximum spam production, wherein the spam checks implemented should be stringent. Overall, we believe this project will help protect the interests of the users of these OSNs and help create a safer community across the web.

- In this project, some parts, such as the topic modeling, will be performed with manual (human) effort. Our future work involves automating this process entirely and producing more robust ways of detecting and banning potentially spam producing posts. Here, we would also strive to make the algorithm more robust by reducing false positives, which can be very harmful if they end up blocking legitimate posts under the false impression that they are spam.

References

[1] Wei, W., Joseph, K., Liu, H. and Carley, K., 2016. Exploring characteristics of suspended users and network stability on Twitter. Social Network Analysis and Mining, 6(1).

[2] Gao H, Hu J, Wilson C, Li Z, Chen Y, Zhao BY. Detecting and characterizing social spam campaigns. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement 2010 Nov 1 (pp. 35-47).

[3] Blog.appriver.com. 2020. Blog | Appriver. Accessed on: July 13, 2020. [Online]. Available at: http://blog.appriver.com/2009/10/zeus-botnettargets-facebook.html.

[4] Abdelmajeed, Nabih. (2019). The Impact of Social Networks on Students' Electronic Privacy in Saudi Arabia Society: Proceedings of the 2018 Computing Conference, Volume 2. 10.1007/978-3-030-01177-2_75.

[5] Gupta, S., Kuchhal, D., Gupta, P., Ahamad, M., Gupta, M. and Kumaraguru, P., 2018, May. Under the Shadow of Sunshine: Characterizing Spam Campaigns Abusing Phone Numbers Across Online Social Networks. In Proceedings of the 10th ACM Conference on Web Science (pp. 67-76).

[6] Wang, D., Irani, D. and Pu, C., 2011, September. A social-spam detection framework. In Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (pp. 46-54).

[7] PyPI. 2020. Pycountry, July 3, 2020. Accessed on: July 15, 2020. [Online] Available: https://pypi.org/project/pycountry/.

[8] Chu, Z., Widjaja, I. and Wang, H., 2012, June. Detecting social spam campaigns on twitter. In International Conference on Applied Cryptography and Network Security (pp. 455-472). Springer, Berlin, Heidelberg.

[9] R.Deepa, Lakshmi & N.Radha,. (2010). Supervised Learning Approach for Spam Classification Analysis using Data Mining Tools. International Journal on Computer Science and Engineering.

[10] E. Ma, Combing LDA and Word Embeddings for topic modeling, Sept. 15, 2018. Accessed on: July 15, 2020. [Online]. Available: https://towardsdatascience.com/combing-lda-and-word-embeddings-for-topic-modeling-fe4a1315a5b4.

Q and A?

?