Text Mining Approaches for Social Media Censoring: A Systematic Literature Review

S.Arthika
Department of Computing and Information Systems
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka
arthika19@gmail.com

K.M.S. Kulathunga
Interac Kanto and North Central Co. Ltd.,
Koushinetsu Branch, GINZA SIX, 6-10-1 Ginza,
Chuo-ku Tokyo, Japan
sanjeewaka19@gmail.com

J. Charles
Department of Physical Sciences and Technology
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka
jspcharles@gmail.com

L.S. Lekamge
Department of Computing and Information Systems
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka
slekamge@appsc.sab.ac.lk

***Abstract* — Social media has recently become one of the most popular communication media used by people all over the world. Their feedback, reviews, and criticism towards products, services, people, or organizations can be easily expressed as posts and comments within the social media platforms e.g., Facebook and Twitter. This mapping study focuses on Social Media Censoring using the text mining techniques which is based on the existing research done in the past eight years. At the starting level of the research, the selected strings which are relevant were applied in major seven digital repositories. Direct search and snowballing on those publications were conducted to scale back the limitation of accessing specific databases. The selected 13 studies have various aspects such as publication year and source, objective, research type, type of methodology used for social media censorship; and those were concerned in this mapping study. From these studies the major area of knowledge regarding social media censorship was gained. The censorship takes place in social media when information, speech, public communication, public posts through images or texts are considered as objectionable, sensitive, or inconvenient for individuals, organizations, or the government. For the purpose of censorship of social media, there are several kinds of text mining techniques available and applied in the real world such as clustering, categorization, and decision tree categorization. The knowledge gained from those studies as an outcome supports building a mechanism that helps to reduce the spread of irrelevant information in social media with the use of text mining techniques.**

Keywords — Social Media, Censorship, Text Mining

I. INRODUCTION

Social media has become the most common way of communication as most of the people in today's world use them in their routine life to keep connected to each other [1]. These social media were introduced at the beginning of the 21st century. In 2002 Friendster was started, MySpace and LinkedIn in 2003, Facebook in 2004, Twitter in 2006 and many others during and thereafter [2]. While this rapid increment of communication sources enhanced easy communication and connection, it also led easier dissemination of false news and information. This situation poses a serious risk to the truthfulness of all media publishing [2]. Censorship is known as omitting the harmful or objectionable information spreading through the media as suppression of speech and public communications. There are mainly four types of censorship; withholding information, destroying information, altering or using selective information, and self-censorship [3].

Nowadays the information shared through the internet is stored in digital repositories. And those repositories contain a large amount of data flows. As per the findings "Text Mining" is an interesting process for the information extraction and to explore knowledge from textual data sources that are located in the digital repositories [4]. According to the researchers, 80% of available texts on the web are unstructured and the remaining 20% is structured [1]. To read those unstructured data there are several kinds of text mining approaches available such as summarization, classification, clustering [4].

Numerous situations across the world highlight the importance of censoring information within social media. Therefore, the traditional methods of censoring are difficult to be executed, warranting automated ways such as the use of text mining techniques.

The research questions attempted to be answered through this mapping study are given below; when and where the related studies regarding social media censorship are published? What are the aspects that have been focused on in the selected study areas? What kinds of research have been done so far? What are the benefits of social media censorship? What are the technologies and methods used in this research area? What are the limitations in existing research?. Presenting answers for the above questions, this mapping study helps the future researchers who are focusing on social media censoring with the aid of text mining techniques while a detailed view on the related research conducted so far in this field of study.

II. OBJECTIVES

The main objective of this mapping study was to investigate the text mining techniques adopted worldwide for social media censoring/ content moderation. In order to achieve the above objective the following specific objectives were set Reviewing the history of social media censorship and its present situation across the world, reviewing different text mining approaches used in social media censorship, providing insights on the future of social media censorship, identifying the barriers and challenges in applying text mining for social media censoring, providing a set of

ISSN 2756-9160 / November 2020.
International Conference on Advances in Computing and Technology (ICACT–2020) Proceedings

guidelines/recommendations for successful implementation of text mining based approaches for social media censoring, and developing a preliminary model for sentiment prediction with the use of social media data.

III. METHODOLOGY

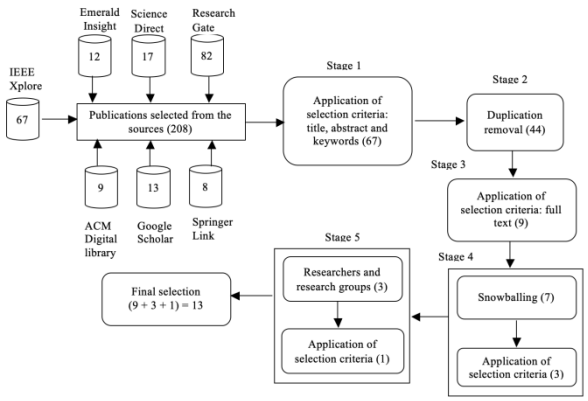

Figure 7: Overall selection process

From the digital repositories, publications during the period from 2011 to 2018 were considered. In the initial search 208 publications were retrieved as a result. The total retrieved results included 67 from IEEE Xplore, 8 from Springer Link, 82 from Research gate, 17 from Science Direct, 13 from Google Scholar, 12 from Emerald Insight, and 9 from ACM Digital Library. Then the criteria of the selection process were continued on the selected publications to identify and extract the most specific related studies. Figure 1 show the major five steps used in selecting the final set of studies.

In the first stage 141 publications were eliminated returning 67 studies. Approximately 67.8% of the studies were eliminated from the selected whole. In the second stage23 duplications were eliminated from the selected 67 publications which were existing in more than one database.

Accordingly, 44 studies were considered under the third stage of the selection process. In this stage full text of the study was considered and incomplete or irrelevant studies were removed. Under this process 35 studies were eliminated considering some specific inclusion criteria (IC) and exclusion criteria (EC). 11 publications were eliminated by EC2 (The paper is published just as an abstract); 11 publications were eliminated by EC5 (The paper is not a primary study. It is either editorial or summary); and finally 13 publications were eliminated because of not satisfying IC1 (The paper describes the text mining techniques for social media censorship).

In the fourth stage of the selection process, the selection criteria focused on outside of the selected sources. The approach helped to overcome the limitations of accessing only specific databases. All the references of the selected 9 publications were analyzed and as a result 7 additional publications were discovered by the method of snowballing. Then the selection criteria were performed throughout these selected 7 studies. Under that 4 publications were selected, which are most relevant to the study. Altogether there were 13 publications selected until then and they were taken into the final fifth stage. From this step 3, publications were identified by direct searching which are mentioned by research of the selected 13 studies. A total of 13 publications were retrieved as a result of this complex selection process.

IV. RESULTS AND DISCUSSION

The findings of this mapping study were summarized in a tabular form having the fields: Study, Aspects, Text Mining Approaches, Solutions and the Limitations/Future Works. This summary table was used as the source of extracting information to answer each research question. Some sample entries of the summary table are given in Table1 representing some of the key studies. Figure 2 elaborates on the distribution of selected papers over the year of publication. Only the papers published from 2011 to 2017 were taken into consideration in order to get an updated view of the study area. Figure 2 further shows the distribution of the selected papers over the digital repository each was obtained from.

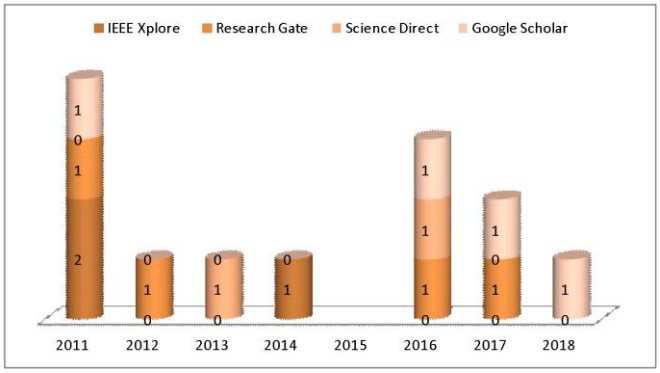

Figure 8: Distribution of the selected studies according to the datasets over the years

The findings revealed the use of some text mining techniques that are applied in censoring purposes such as Information Extraction, Categorization, Clustering, Topic Tracking, Summarization and Association Rule Extraction [1], [5]. Moreover, various limitations associated with the above studies were identified while identifying future research directions to facilitate social media censoring with text mining techniques in order to provide better solutions. Following are among the findings.

1. The limitation associated with the selection of areas for censorship experiments.
2. Limitations associated with the social media selected due to the restrictions within the country.
3. While text data is extensively being used currently in isolation, image, audio, and video data can be successfully integrated in the future.

ISSN 2756-9160 / November 2020.
International Conference on Advances in Computing and Technology (ICACT–2020) Proceedings

Table 3: Partial Summary Table

Study	Aspects	Text Mining Approaches	Solutions	Limitations/ Future Works
[1]	N/A	IE Categorization Clustering Topic Tracking Summarization Question Answering Association Detection	Need of text mining based DSS for government agencies. Various text mining applications developed in e-government. Proposed an integrated framework used by government organizations' to develop text mining based DSS.	Extend the technical architecture for Text-Mining based DSS for e-governance as Multi-lingual Text Mining based Decision support system (MLTMbDSS)
[3]	Facebook and Twitter	NLP Text Clustering Categorization Association Rule Extraction	Categorization of text mining as; text clustering, text categorization, association rule extraction and trend analysis were according to applications.	Examining the text mining techniques on Arabic textual data from Facebook and Twitter, & sentiment analysis of Arabic text.
[6]	Twitter	Topic Extraction Clustering	Found the vast bulk of censored tweets contained political content, often critical of the Turkish government. Also the research establishes that Twitter radically under-reports censored tweets in Turkey, raising the possibility of similar trends hold for censored tweets from other countries. And has discussed the relative ease of working around Twitter's censorship mechanisms.	The research conducted within the Twitter data collected targeting only Turkey with some Twitter rate limitations and some restrictions from other groups those who expect to crawl their content.
[7]	Social Media	N/A	The censorship of expression that exists on social media in China shows major setbacks in regard to the key elements of a public sphere	Focused only on the concept of the public sphere but also can analyze social media and their impact on freedom of expression on democracy and ideology as well.

(N/A – Not Available)

4. Text data from social media posts and comments can be supplemented by articles from newspapers and direct interviews as a means of enhancing the prediction accuracy.

V. CONCLUSION

The main aim of this mapping study was to provide a comprehensive idea on social media censorship with the use of text mining techniques which overcome the barriers and limitations in conventional manual mechanisms. The study revealed the use of various text mining approaches that can be successfully adopted in social media censoring. Further, the limitations and barriers associated with the existing works were also identified while providing insights on prospective research directions which enhance the existing mechanisms of text mining techniques. While sentiment analysis, hate speech detection, and fake news identification are among the frequent text mining tasks performed on social media data, the development of a prediction model and comparison among different machine learning algorithms including Logistic Regression, Support Vector Machine, and Random Forest are looked forward in the study with related to one of the above tasks. It is believed that the findings of the present mapping study would contribute and support various attempts across the world for assuring a reliable social media experience

REFERENCES

[1] K. Rao and S. Dey, "Decision Support for E-Governance: A Text Mining Approach," *International Journal of Managing Information Technology*, vol. 3, no. 3, pp. 73–91, 2011.

[2] T. Mayeesha, Z. Tasneem, J. Jones, and N. Ahmed, "Applying Text Mining to Protest Stories as Voice against Media Censorship," 2018.

[3] S. A. Salloum , et al., "Mining Social Media Text: Extracting Knowledge from Facebook," Int. J. Comput. Digit. Syst., vol. 6, no. 2, pp. 73–81, 2017, doi: 10.12785/ijcds/060203.

[4] R. Chandramouli, "Emerging social media threats: Technology and policy perspectives," 2011 2nd Worldw. Cybersecurity Summit, WCS 2011, 2011.

[5] S. A. Salloum, M. Al-Emran, A. A. Monem, and K. Shaalan, "A survey of text mining in social media: Facebook and Twitter perspectives," Adv. Sci. Technol. Eng. Syst., vol. 2, no. 1, pp. 127–133, 2017, doi: 10.25046/aj020115.

[6] R. S. Tanash, Z. Chen, T. Thakur, D. S. Wallach, and D. Subramanian, "Known unknowns: An analysis of twitter censorship in Turkey," WPES 2015 - Proc. 2015 ACM Work. Priv. Electron. Soc. co-located with CCS 2015, pp. 11–20, 2015, doi: 10.1145/2808138.2808147.

[7] J. Power, Testing Habermas ' ideas on the public sphere on social media in China. 2016.