

## Social Media Analytics – Challenges during the Data Analysis

Samta Jain Goyal<sup>1</sup>, Rajeev Goyal<sup>2</sup>  
Department of Computer Science and Engineering  
[1sjgoyal@gwa.amity.edu](mailto:1sjgoyal@gwa.amity.edu), [2rgoyal@gwa.amity.edu](mailto:2rgoyal@gwa.amity.edu)  
Amity University, Gwalior, MP, India

### Abstract

Social media in the present time is being analyzed due to many perspectives. The analysis process contains four basic steps so that it can study in deep and put accordingly. These four steps are ordered wise i.e. first we need to discover the data known as “Data Discovery” and then “Data Collection”, “Data Preparation” and last but not the least is “Data Analysis”. There are great challenges found during the “data analysis phase”. For the present scenario, seriously need to focus on the data means what kind of data is used, what purpose to float data, what it means, what is the intension of the data. To get the answer to those, need a structured framework for the analysis. Also, the research findings are used to help in the analysis part which is quite challenging.

**Keywords:** Data Analysis, Data Discovery, Social Media Analytics, Information systems.

### 1. Introduction

Over the last decades, Social Media playing a very enormous role to spread information, knowledge in different segments such as in politics, social networking sites, business, entertainment, science, etc. The main reason to float our knowledge, our information is the nature of human being and second most important things are spreading cost which is quite low now a day as compared to earlier. Sharing information, communication, put our thoughts and ideas are giving the enormous growth in social media, these data on social media where the huge amount of data entertain is called “Big-Data”. On Social media, different types of data, we entertain such as types of data like audio, text, videos, geolocations, pictures, etc. Whatever data we upload or download or use are categorized into two main sections. One is structure Data and the other is non-structured data. On social sites, all textual data comes under the category of unstructured data and the rest of the data comes under the “Structured Data” category. There are new opportunities in the market to find a pattern of data based on the analysis of the social media's contents. So that we can find the intention, motto, objectives of the statements or comments. In 2011, Golder,

and Macy found after the analysis that Twitter data generally shows the mood of the person which swings based on the day and season. Social Media data plays a critical role in Information System which are used to find the influenced factor of the network (Susarla, Oh, & Tan, 2012).

In fact, many research papers shown the case studies where a large amount of data is taken, find out a pattern based on the specific time, specific subject, specific disciplines, etc. The objective of the “Social Media Analytics” is to combine, extend, and adapt some methods and approaches to the study of data into a deep manner so that we can easily identify the purpose or intension of uploaded data of any form. In 2013, Aral et. al has given a framework that is used to organize social media research. The same thing is done by Osch and Coursaris in 2013 with some new ideas. During research analysis, uses specific methods to analyze social media contents. Social Media Analytics has four steps in order wise. These are “Data discovered”, “Data collection” and “Data preparation” and “Data Analysis”. “Data Analysis” phase is demanded in the present scenario due to huge amount of

contents put onto the social media. So, this is a challenge in “Social Media Analytics” which is used to manage the complexity of conducting data analysis. This paper is organized in the following sections where the first section is used to provide the status of the Literature for the “Social Media Analytics”, also focus the theoretical background of this paper. In the second section, describing the research design and highlights the research findings. The third section of this paper will show the results of the proposed idea of the model for the “Social Media Analytics” and their impact. At last, a conclusion with future aspects is presented.

## **2. Literature Review**

In the field of Social Media Analytics, we need to deal with the approaches which are used to analyze the social media contents. These contents are frequently uploaded. Social Media Analytics contains discovery, collection, preparation, and analysis steps described by Stieglitz et al. in 2014.

### **a. Social Media Analytics**

As the name of the Social Media which means that it's open to society and support many application areas. Even then some of the features are common due to the huge amount of content uploaded on daily basis through many users, organizations.

In current times, the analysis of social media contents becomes the main research area in Information Systems. Here, the main objective is to find and analyze the information diffusion given by Liu in 2015 and by Zhang & Zhang in 2016. Social media contents are used to detect new trends or issues with existing trends. Social media platforms provide a different source of data for the journalism & communication fields especially in politics where the Political parties and governments use social media as a platform to communicate with users. Overall we say that Social Media Analytics is a very complex process because it handles and processes various types of data and steps to get some relevant results. These processes follow some standardized framework and process model.

### **b. Phases in Social Media Analytics**

Researchers have been developed the framework for Social Media Analytics. In 2013, Aral et al. described

the research opportunities related to the Social Media Analytics. And designed a framework that has shown the relationship between the social –media, society and business. This framework has multiple levels and different types of social media-related activities. In 2013, Osch and Courser have specified their studies based on the actor, artifact and the activities they examined. In 2014, Fan and Gordon proposed a method for Social Media Analytics which consists of three steps. These steps are “capture”, “understand”, and “present”. Same in the year 2014, Stieglitz et al. also proposed a framework for Social Media Analytics (SMA). This design is approved by many of the Information Systems. There still are many challenges where researchers need to work such as data sources, approaches, software architecture and data storage.

In any standard designed framework, the Discovery step is used to uncover the structures and any specified patterns. The tracking step that framework includes the decisions on the data source. In the preparation step, the original framework used to prepare data for the last step of the framework. The last step is an analysis where we use several methods which include mining and pattern related techniques.

## **3. Challenges in Big-Data analytics**

“Big-Data” on social media includes different types, different sources and different disciplines of data. Social media data generally include Big-Data which is also known as a “Social Big-Data” (Guellil & Boukhalifa, 2015) or “Social Media Big-Data” (Lynn et al., 2015). Big-data and traditional data have a huge difference because of its volume, velocity, variety and of course its veracity which shows the uncertainty concerning its quality. This veracity feature uses by the researchers in security-related issues, especially in authenticity and integrity.

To design a new framework, a variety of data generally influence the structure. Big-Data generally concerns with the ethics and accessing barriers in research. Also, lack of data and content accuracy, need to be chosen the accurate and effective methods for the data extractions. These types of quality concerns come under the category

of Veracity. So in big-data concepts, it is the main criterion for the assessment. For Social Media Analytics, the classification of data through the four V's is quite important. The Four V's of Big Data are volume, variety, velocity and veracity. but at the same time, the classification phase is very important because this phase is used to determine the type of data on social sites.

#### 4. Proposed Research design

Social Media Analytics is a developing field of the research, which is based on logical conceptualization. Our proposed design contains two main and basic steps, shown in Figure-1.

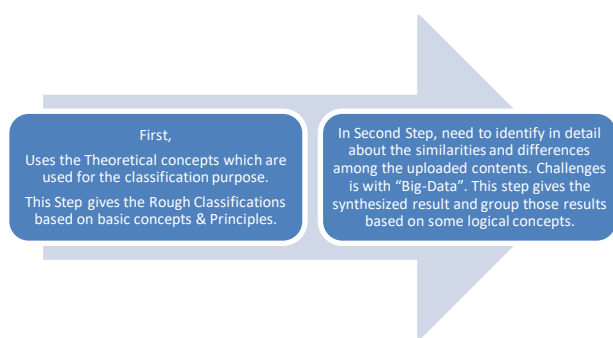


Figure-1: Proposed design framework

In general search or analysis for the uploading articles is first to need to search some predefined terms in the selected datasets and then read its title to determine its relevance. But the main issue occurs during the Big-Data where need to mention or design some formulas or functions which will reduce the time of searching for the analysis. This thing is also reducing the irrelevance of the uploaded articles. Social Media Analytics is specifically prominent for the datasets. For the potential relevance of the search, the discovery, collection and/or preparation phases must work independently so that the Data Analysis phase can work more effectively. We should also use the backward search technique for the feedback or find confirmed output. During all these steps, just need to specify criteria of the contexts which generally cited frequently.

#### 5. Research Findings

The systematic execution for searching through the search terms with all the possible combinations of "AND" and "OR" in the predefined data sets. During the process, search query, relevant with the specific methods for the extraction, classification based on the concept of

machine learning, gives the more effective output. This analysis result of the data decides the category of the contents.

For a huge amount of data, there is a challenge in the discovery step. There is work for the discovery in data. There we need to work with the high-velocity, high-variety, and the low-veracity environment. For the classification, all the articles should group under the common heading. Specifically, when they face the same challenges. So obviously they need a more detailed analysis of the classification. Social Media Analytics is an interdisciplinary field (Stieglitz et al., 2014). So, the researchers are working on different backgrounds and different disciplines. Every background and disciplines have their own merits and demerits. Being researchers need to apply some computational approaches which give benefits to the existing social theory.

Qualitative analysis is used for extraction and classification purposes. This also allows an in-depth qualitative content analysis. Sometimes the volume of data makes things so difficult. So, the need for a structured and efficient approach is high across all of the application areas of Social Media Analytics. For event detection on social media, use modeling algorithms with some specific characteristics and unique challenges.

#### 6. Discussion

In this paper, we have shown the objectives, challenges which researchers are faced during the Social Media Analysis. The steps concerned with that analytics are quite not easy for processing. There are plenty of challenges in each level of the work, and they all need to be addressed properly to get success in this work. There we all as researchers need to pay more attention to each level of the analysis in terms to identify the type of data, the meaning of the data, the intention of the data. The four V's approach of the Big-Data Literature shown the biggest challenge related to the data. Specifically, with the "Volume of the Data". All of the data uploaded are not relevant enough so filtering the data and detection of relevant data can have done through some advanced detection algorithms. On the next stage, play with the types of data. Once social media entertains varieties of data, which again a complicated task for the collection

and finds its types. So, need to identify some sophisticated software architectures.

## 7. Conclusion

Social Media Analytics is new research area as well as an interesting too because there we can find type of data, its intension, meaning to post that data etc. So there is a vast scope to develop new and effective system for the identification and understanding the social media. This paper is designed for the presenting basic overview, the objective of the Social Media Analytics, related challenges during the analysis of the contents. It was also found the conceptual process where discovery, collection, preparation, and analysis phases are required to fulfil the base requirement of the analysis. Also, this paper presents some solutions to the challenges which were relevant to the research.

## References

- [1] Abbasi, A., Fu, T., Zeng, D., & Adjeroh, D. (2013). Crawling credible online medical sentiments for social intelligence. *Proceedings – SocialCom/PASSAT/BigData/EconCom/BioMedCom* 2013, 254–263. <http://dx.doi.org/10.1109/SocialCom.2013.43>.
- [2] Al-Qurishi, M., Al-Rakhami, M., Alrubaian, M., Alarifi, A., Rahman, S. M. M., & Alamri, A. (2015). Selecting the best open source tools for collecting and visualizing social media content. 2015 2nd world symposium on web applications and networking (WSWAN), 1–6. <http://dx.doi.org/10.1109/WSWAN.2015.7210346>.
- [3] Aral, S., Dellarocas, C., & Godes, D. (2013). Introduction to the Special Issue—Social media and business transformation: a framework for research. *Information Systems Research*, 24(1), 3–13. <http://dx.doi.org/10.1287/isre.1120.0470>.
- [4] Beier, M., & Wagner, K. (2016). Social media adoption: barriers to the strategic use of social media in SMEs. *Proceedings of the European conference on information systems*.
- [5] Chen, F., & Zhang, L. (2016). How to integrate social media in the IS curriculum, especially for a small IS program? *Proceedings of the Americas conference on information systems*.
- [6] Chen, S., Guo, C., Yuan, X., Zhang, J., & Zhang, X. L. (2014). MovementFinder: Visual analytics of origin-destination patterns from geo-tagged social media. 2014 IEEE conference on visual analytics science and technology, VAST 2014 – proceedings, 239–240. <http://dx.doi.org/10.1109/VAST.2014.7042509>.
- [7] Driscoll, K., & Walker, S. (2014). Working within a black box: Transparency in the collection and production of big twitter data. *International Journal of Communication*, 8(1), 1745–1764.
- [8] Griffiths, M., & McLean, R. (2015). Unleashing corporate communications via social media: A UK study of brand management and conversations with customers. *Journal of Customer Behavior*, 14(2), 147–162. <http://dx.doi.org/10.1362/147539215X14373846805789>.
- [9] Hiltz, S.R., Diaz, P., & Mark, G. (2011). Social media and collaborative systems for crisis management, *ACM Transactions on Computer-Human Interaction*, Vol. 18 Issue 4, December 2011, ACM New York, NY, USA, DOI: <https://doi.org/10.1145/2063231>.
- [10] S. Stieglitz et al. *International Journal of Information Management* 39 (2018) 156–168 166 2063232.
- [11] Jalonen, H., & Jussila, J. (2016). Developing a conceptual model for the relationship between social media behavior, negative consumer emotions, and brand disloyalty. *Lecture Notes in Computer Science*, 9844, 134–145. [http://dx.doi.org/10.1007/978-3-319-45234-0\\_13](http://dx.doi.org/10.1007/978-3-319-45234-0_13).
- [12] Jungherr, A., Schoen, H., & Jürgens, P. (2016). The Mediation of Politics through Twitter: An Analysis of Messages posted during the Campaign for the German Federal Election 2013. *Journal of Computer-Mediated Communication*, 21(1), 50–68. <http://dx.doi.org/10.1111/jcc4.12143>.
- [13] Kefi, H., Mlaiki, A., & Kalika, M. (2015). Social Networking Continuance: When Habit leads to information overload. *ECIS 2015 Proceedings*, 1–13.
- [14] Kleindienst, D., Pflieger, R., & Schoch, M. (2015). The business alignment of social media analytics. *ECIS 2015 Completed Research Papers*, 4801, Paper 103.
- [15] Li, Z., & Huang, K.-W. (2014). The Monetary Value of Twitter Followers: Evidence from NBA Players. In *Proceedings of the International Conference on Information Systems*.
- [16] Moalla, I., Nabli, A., Bouzguenda, L., & Hammami, M. (2017). Data warehouse design approaches from social media: Review and comparison. In *Social Network Analysis and Mining (Vol. 7)*. <https://doi.org/10.1007/s13278-017-0423-8>.
- [17] Rehman, N. U., Weiler, A., & Scholl, M. H. (2013). OLAPing social media: The case of Twitter. *International Conference on Advances in Social Networks Analysis and Mining*, 1139–1146. <http://dx.doi.org/10.1145/2492517.2500273>.
- [18] Zeng, D., Chen, H., Lusch, R., & Li, S. H. (2010). Social media analytics and intelligence. *IEEE Intelligent Systems*, 25(6), 13–16. <http://dx.doi.org/10.1109/MIS.2010.151>.
- [19] Zhang, H., Sun, Z., Liu, Z., Xu, C., Wang, L. (2015). Dart: A Geographic Information System on Hadoop. In *Proceedings – 2015 IEEE 8th International Conference on Cloud Computing, CLOUD 2015* 90-97. <https://doi.org/10.1109/CLOUD.2015.22>.
- [20] Zhang, S., Zhao, L., Lu, Y., & Yang, J. (2016). Do you get tired of socializing? An empirical explanation of discontinuous usage behavior in social network services. *Information & Management*. Advance online publication. <https://doi.org/10.1016/j.im.2016.03.006>.
- [21] Zhang, Z., & Zhang, Y. (2016). How Do Explicitly Expressed Emotions Influence Interpersonal Communication and Information Dissemination? a Field Study of Emoji? S Effects on Commenting and Retweeting on a Microblog Platform. In *PACIS 2016 Proceedings*.
- [22] Zhao, Z., Feng, Z., Zhang, Y., Ning, L., Fan, J., & Feng, S. (2015). Collecting Managing and Analyzing Social Networking Data Effectively. In *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)* 1642–1646. <https://doi.org/10.1109/FSKD.2015.7382191>.