

# Exploring Data Mining Techniques in Social Media Networks

Narendra Kumar Verma<sup>1\*</sup> Dr. Prabhat Pandey<sup>2</sup>

<sup>1</sup> Research Scholar

**Abstract** – Today, the utilization of social networks is developing incessantly and quickly. Additional disturbing is the way that these networks have turned into a generous pool for unstructured data that have a place with a host of domains, including business, governments and health. Data mining is the strategy of examining prior databases to create new helpful information by utilizing a few methodologies and executing specific operations. The goal of the present survey is to break down the data mining procedures that were used by social media networks. Social Media in the most recent decade has increased momentous attention. This is ascribed to the moderateness of getting to social network sites, for example, Twitter, Google+, Facebook and other social network sites through the internet and the web 2.0 technologies. Numerous individuals are getting to be keen on and depending on the social media for information and supposition of different clients on various topics.

**Keywords:** Data, Mining, Social, Media, Network, Techniques, etc.

-----X-----

## 1. INTRODUCTION

Data mining alludes to removing or mining knowledge from a lot of data. The term is really a misnomer. Accordingly, data mining ought to have been more fittingly named as knowledge mining which emphasis on mining from a lot of data. It is the computational procedure of finding designs in huge data sets including methods at the convergence of man-made reasoning, machine learning, insights, and database systems. The general goal of the data mining procedure is to separate information from a data set and transform it into a justifiable structure for further use.

The key properties of data mining are-

- Automatic disclosure of examples
- Prediction of likely outcomes
- Creation of significant information
- Focus on huge datasets and databases

A social network is a set of individuals or associations or other social elements associated by set of social connections, for example, kinship, cooperating or information trade. Social network examination centers on the investigation of the example of connections among individuals, associations, states and such social substances. In this paper a survey of the works done in the field of

social network investigation is done and this paper likewise focuses on the future patterns in research on social network examination.

• Social media network as online services that enable people to build an open or semi-open profile inside a limited system, articulate a rundown of different clients with whom they share an association, and view and navigate their rundown of associations and those made by others inside the system. Data mining systems have been observed to be equipped for taking care of the three dominant debates with social network data namely; size, noise and dynamism. The voluminous idea of social network datasets require automated information handling for dissecting it inside a sensible time. Strikingly, data mining procedures likewise require gigantic data sets to mine remarkable examples from data; social network sites have all the earmarks of being ideal sites to mine with data mining tools. This forms an empowering factor for cutting edge indexed lists in web crawlers and furthermore helps in better comprehension of social data for research and hierarchical capacities. Data mining tools surveyed in this paper ranges from unsupervised, semi-supervised to supervised learning.

### 1.1 Social Media Background

Social Media assume a functioning job in publicizing how individuals feel about specific products/services, issues and occasions in many

parts of life. This can be said to be because of the reasonableness of getting to the social network sites through the internet and the advances in web 2.0 technologies. More individuals are becoming inspired by and depending on the SM for information, breaking news and other differing topics. Clients discover what other individuals' perspectives are about sure item/administration, film, school, or significantly more major issues like looking for other individuals' supposition on political hopefuls in national race survey .Millions of individuals get to SM sites, for example, Twitter, Facebook, LinkedIn, YouTube and MySpace to scan out for information, breaking news and news refreshes.

## 1.2 Data Mining Techniques Used in Social Media

Data mining strategies are fit for taking care of the three dominant debates with SM data which are size, noise and dynamism. SM data sets are extremely voluminous and require automated information handling for examining it inside a sensible time. As data mining additionally require colossal data sets to mine remarkable examples from data, SM sites give off an impression of being ideal sites to chip away at particularly where feeling/sentiment articulation is included:

- **Unsupervised Classification**

This can be by method for uncovering phrases that incorporate descriptive word or modifiers (grammatical feature tagging).The semantic direction of each phrase can be approximated utilizing PMI-IR and after that arrange the audit utilizing the normal semantic direction of the phrase. Cogency of title, body and comments created from blog entry has likewise been utilized in bunching similar web journals into huge gatherings. For this situation catchphrases assumed important job which may be multifaceted and exposed.

- **Supervised Classification**

While bunching methods are utilized where premise of data is built up however data example is obscure, grouping procedures are supervised learning systems utilized where the data association is as of now recognized. Pre-preparing and considering security privileges of individual (as mentioned under research issues of this paper) ought to likewise be considered. In any case, since SM is a dynamic platform, impact of time must be balanced in the issue of theme acknowledgment, however not significant on account of network enlargement, bunch conduct/impact or marketing. This is on the grounds that this ascribes will undoubtedly change from time to time. Information refreshes in some SM, for example, twitters and Facebook present.

- **Semi-supervised Classification**

Semi-supervised learning is a goal-focused on action yet not at all like unsupervised; it tends to be explicitly assessed. Creators took a shot at a mini training set of seed in positive and negative articulations chose for training a term classifier. Synonym and antonym comparatives were added to the seed sets in an online lexicon. The methodology was meant to create the all-encompassing sets 'P' and 'N' that makes up the training sets.

## 2. LITERATURE REVIEW

**Thiel K, et al (2012)** expressed that social media gives an abundance of chances to taking advantage of client inclinations, assessments, opinions, surveys and evaluations about substance and products, and that, this training has been progressively prominent among the corporate companies. There is an outsized amount of data accessible because of the enormous assortment of social media diverts in this day and age of social media.

**Gundecha P and Liu H (2012)** proposed that sentiment investigation and opinion mining permit the business to chip away at the development estimates of item sentiments, brand perception, new item perception, and notoriety management.

**Chandrakala and Sindhu (2012)**, talked about that there is a need to investigate, examine and sort out the client's perspectives, input and suggestions as opinions. Opinion Mining utilizes Natural Language Processing and Information Extraction to distinguish the opinions and perspectives on the client and characterize the documents by point.

**Wu He et al (2013)**, uncovered the estimation of social media competitive investigation and the intensity of content mining as a successful system to extricate business esteem from the immense amount of accessible social media data.

## OBJECTIVES

1. To explore Data Mining Techniques.
2. To explore about Social Media.
3. To analyze social media networks.
4. To synthesize the data extracted from the selected articles.
5. To know the selected article type's distribution.

## RESEARCH METHODOLOGY

In this stage, we investigated the articles chose to separate the information required to address the

exploration questions. In this way, we have structured an extraction form to remove the required data. In view of the extraction form, two creators assumed the job of extraction and checking. If there should arise an occurrence of a disagreement between the extractor and checker, bunch meetings were directed between all creators to determine any issue. Some challenges happened during the extraction procedure.

In this examination we will follow set of methodology to do this work:

- **Data Collection:** - Data will be gathered from any social network sites.
- **Data preprocessing:** Data gathered from social media will be changed over from unstructured to organized data set. Likewise missing, excess, invalid qualities and so on will be removed.
- **Data integration:** - More than one data set which has same quality will be incorporated for successful knowledge forecast.
- **Data mining tool implementation and model testing:** - Available data mining tool will be utilized for implementation of our proposed model; exactness testing will likewise be spread in this progression.
- **Knowledge Production and basic leadership:** - Based on knowledge created by our proposed model, would be used for basic leadership

**ANALYSIS AND DISCUSSION**

**Table 1: Data Extraction Form**

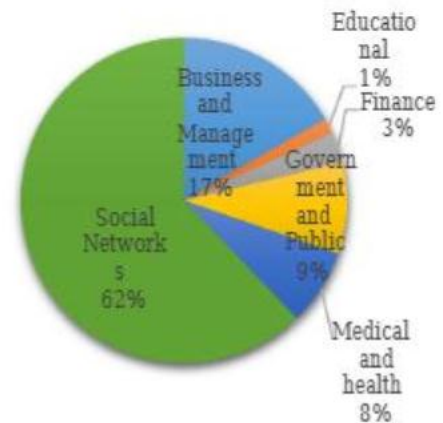
Article ID
Data Extractor
Data Checker
Publication Year
Authors
Article Source
Article Title
Article Type
Domain
RQ 1
RQ 2
RQ 3
RQ 4
RQ 5

To synthesize the data extricated from the chose articles, we utilized various procedures to aggregate evidence that will answer the RQs. The accompanying clarifies the synthesis procedure we followed in detail:

We utilized the story synthesis method were the extricated information was organized. For the data removed (quantitative) ,which came from various articles that have different exactness count strategies, we utilized paired outcomes to measure the outcomes, which are demonstrated in a comparable way. The qualities and shortcomings of the data mining systems have the same meaning however are written in various ways. Thusly, to bind together these focuses, we pursued the corresponding translation method which is considered as one of the systems that can be utilized for blending the subjective data.

**Table 2: Selected Article Type’s Distribution**

Article Type	Freq.
Case Study	4
Experiment	60
Survey	2
<b>Grand Total</b>	<b>66</b>



**Figure 1: Domain among Articles**

MohammadNoor Injadat, Fadi Salo and Ali Bou Nassif, Data Mining Techniques in Social Media: A

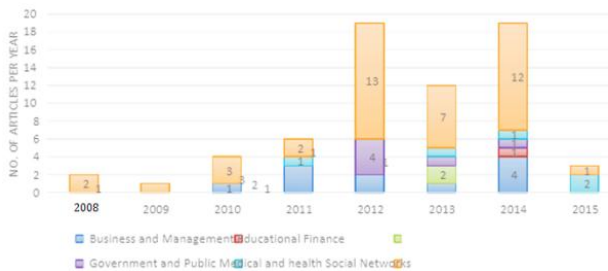
Survey, Neurocomputing, <http://dx.doi.org/10.1016/j.neucom.2016.06.045>

From the chose articles, we recognized six general domains which connected different procedures in nine diverse research regions to mine the progression of enormous data accumulated from social media. The rundown of these domains pursues:

- Business and Management (BM)
- Education (EDU)
- Finance (FIN)

- Government and Public (GP)
- Medical and Health (MH)
- Social Networks (SN)

Figure given below demonstrates that social networks and business and management were the most dynamic domains utilized by data mining methods, with a level of 79% among all domains. Government and open with a level of 9% speaks to the third dynamic domain. Index (A), Table 2, incorporates nitty gritty information pretty much all domains.



**Figure-2 Domains distribution per year**

- Semantic analysis
- Sentiment analysis

Figure demonstrates further information about the findings by illustrating the distribution of the domains applying data mining techniques per year. Based on the figure, it can be clearly seen that the number of publications has increased dramatically in 2012 and 2014 with 19 articles in 5 domains for both periods. In 2013, the number went down to 12 articles in 5 domains. The social network data analysis remains the most active domain among the considered period. Among the selected articles, we identified 9 active research objectives adopted data mining techniques.

## CONCLUSION

Various data mining strategies have been utilized in social network examination as canvassed in this survey. The systems go from unsupervised to semi-supervised and supervised learning methods. So far various dimensions of accomplishments have been accomplished either with lone or combined methods. The outcome of the experiments led on social network investigation is accepted to have revealed more insight into the structure and exercises of social network. The various experimental results have additionally confirmed the importance of data mining methods in recovering significant information and substance from enormous data produced on social network. Future survey will in general research novel cutting edge data mining procedures for social network examination. The survey will compare

similar data mining tools and recommend the most appropriate tool(s) for the dataset to be broke down.

## REFERENCES

1. Thiel K, et. al. (2012). "Creating Usable Customer Intelligence from Social Media Data: Network Analytics meets Text Mining", KNIME. pp. 1-18.
2. Gundecha P. & Liu H. (2012). "Mining Social Media: A Brief Introduction", INFORMS. DOI: <http://dx.doi.org/10.1287/educ.1120.0105>. 2012.
3. S Chandrakala, C Sindhu (2012). "Opinion Mining and Sentiment Classification: A survey", ICTACT Journal on Soft Computing. 3(1). pp. 420-427.
4. Wu He, S Zha, L Li. (2013). "Social media competitive analysis and text mining: A case study in the pizza industry", International Journal of Information Management, Elsevier.33. pp. 464-472.
5. Ghosh, R. & Lerman, K. (2011). Parameterized centrality metric for network analysis. Physical Review E, 83(6), 066118.
6. Ruan, X. H., Hu, X., Zhang, X. (2014). Research on Application Model of Semantic Web-Based Social Network Analysis. In Proceedings of the 9th International Symposium on Linear Drives for Industry Applications, Volume 2 (pp. 455-460). Springer Berlin Heidelberg, 2014
7. Pham, M. C., Cao, Y., Klamma, R., Jarke, M. (2011). A clustering approach for collaborative filtering recommendation using social network analysis. J. UCS, 17(4), pp. 583-604.
8. Liu, F., Lee, H. J. (2010). Use of social network information to enhance collaborative filtering performance. Expert Systems with Applications, 37, pp. 4772-4778.
9. Newman, M. (2010). Networks: An introduction. Oxford University Press.
10. Scott, J. (2011). Social network analysis: developments, advances, and prospects. Social network analysis and mining, 1(1), pp. 21-26.

**Corresponding Author**

**Narendra Kumar Verma\***

Research Scholar