



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Issue 4, April 2017

## A Review on Various Data Mining Techniques in Social Media

Dr.B.Umadevi<sup>1</sup>, P.Surya<sup>2</sup>

Assistant Professor & Head, P.G. & Research Department of Computer Science, Raja Doraisingam Govt. Arts College, Sivaganga, Tamil Nadu, India<sup>1</sup>.

Research Scholar, P.G. & Research Department of Computer Science, Raja Doraisingam Govt. Arts College, Sivaganga, Tamil Nadu, India<sup>2</sup>.

**ABSTRACT:** The tremendous changes in the technological world will always make the people to get deeper in to that. In that series today we all are well connected with each other through the social media network. The Face book, WhatsApp is the most popular social network among the society. Anyone who has cellular phone might have a touch with the above said social Medias. Various researchers put their insight in the three major challenging characters of the SM (Social Media) like large, noisy and dynamic. The researcher wants to make analyses on the data which are being used in the social Medias to overcome and control through the data Mining techniques.

**KEY WORDS:** Face book; Social media; Data volume; Clustering; Classification

### I. INTRODUCTION

Social media are defined as virtual spaces where people of all ages can make contacts, share information and ideas, and build a sense of community [1]. An exceptional amount of data is present due to the global use of social media and concentration to various branches of learning such as sociology, business, psychology, politics, news and other educational aspects of societies.

Facebook, twitter, MySpace, and bebo, can be taken as most commonly accessed social media sites. Social media can be used in many business activities like increasing word-of-mouth marketing, marketing research, General marketing, Idea generation & new product development, Co-innovation, Customer service, Public relations, Employee communications and in Reputation management [2].

With a remarkable grow of social media in the recent years, the application of data mining methods to social media data is also getting increasing recognition. The three challenging characteristics of the social media are that it is large, noisy and dynamic. These challenges are overcome by using data mining techniques. The following Fig. 1. explains about social media mining is the process of representing, analyzing, and extracting actionable patterns and trends from [2][17] raw social media data. The term "mining" is an analogy to the resource extraction process of mining for rare minerals. These patterns and trends are of interest to organizations and they can use these patterns and trends to design their strategies or introduce new programs. Social media mining uses a range of basic concepts from computer science, data mining, machine learning and statistics.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Issue 4, April 2017

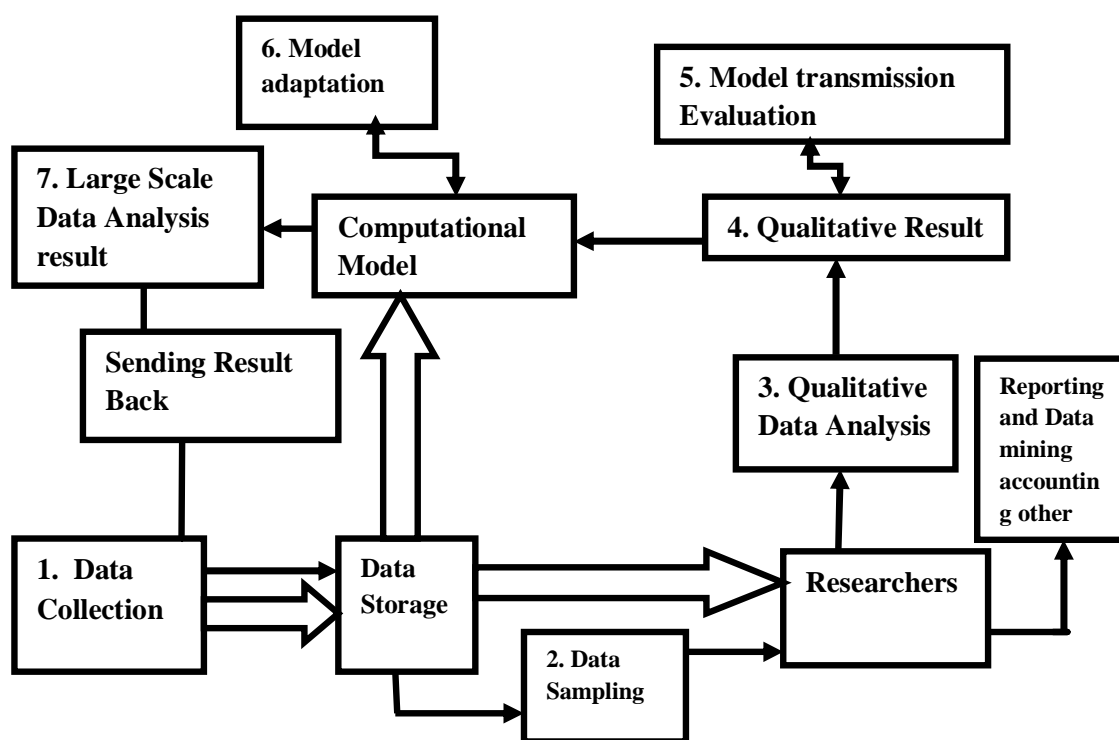


Fig. 1: Workflow of Social Media Data Mining

## II. RELATED WORKS

Raju and Sravanthi have studied the issues related to the social networks using web mining techniques. Their analysis focuses highly on the applications of web mining techniques and a general process for social networks. It mainly highlights the difficulties in selecting data samples, finding communities and patterns in social networks and analyzing the overlapping [3] communities.

Rahman deliberative a systematical data mining approach to mine intellectual knowledge from social data. The author took face book as a primary data source and proposes to use different data mining techniques to analyze the social networking site and other sites. He used K-nearest neighbor (KNN) to classify objects based on [4] samples.

Mosley discussed the application of correlation, clustering, and association analyses to social media. The author describes how data mining and text analytics can be applied to social media in order to identify key themes in the data. The issues in terms of accuracy while collecting the data [5] from social media were also highlighted.

Adedoyin-Olowe, Gaber and Stahl have studied on the techniques that are currently used to analyze SM. It analysis of social media data has proved to be effective, this is so because of the capacity data mining possess in handling noisy, large and dynamic data. According to the authors, in future to mine the data generated on SM currently used and yet-to-be-explored [6] data mining techniques will be used.

Hetherly et.al explained how deduction on attacks in social networking data to foresee undisclosed information of users through various techniques and projected the efficacy of the techniques through dataset obtained from the Dallas/Fort Worth, Texas network of the Face book social networking [7] application.

S.G.S Fernando have studied existing web mining techniques used to mine social network data. The paper has specifically focused on the techniques used to mine social network data. Most of the algorithms are developed to mine the structure of the social network where mapping the network [8] to a graph.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Issue 4, April 2017

Mariam Adedoyin-Olowe et.al discussed different data mining techniques used in mining diverse aspects of the social network over decades going from the historical techniques to the up-to-date models, including our novel technique named *TRCM*. that techniques range from unsupervised to semi-supervised and supervised learning methods. So far various levels of successes have being achieved either with solitary or combined techniques [9]. The outcome of the experiments conducted on social network analysis is believed to have shed more light on the structure and activities of social network.

### III. DATA MINING TECHNIQUES AND THEIR CHARACTERISTICS IN SOCIAL MEDIA

Data mining is the process used to extract usable data from a large set of any raw data [10]. Data mining has many application fields like science, medical, banking and research [11]. Different data mining techniques have been developed by scientists in order to beat the problems such as size, noise and dynamic environment of the social media or social network data. Due to the huge quantity of data in the social media, a routine data processing is required in order to analyze it within a specified time span [12]. Table. 1. which explain the characteristics of social media analysed by a variety of authors by means of data mining algorithms.

**TABLE. 1. DATA MINING TECHNIQUES AND THEIR CHARACTERISTICS IN SOCIAL MEDIA**

S.NO.	ALGORITHM	CHARACTERISTICS	AUTHORS/YEAR
1	Vector Space Model	Mining social networks of person entities using Improved Vector Space Model (IVSM) [13]. Based on the testing that has been done to analyze Social networks, they authors state that their approach is effective.	F. Yang ,Z. Xu, S. Li, X. Li (2010).
2	Graph mining algorithms	Bourqui et.al presented a framework which is based on dynamic graph discretization and graph clustering. This framework is capable of detecting the dynamic changers of the social network structure and identifies events analyzing temporal dimension and exposes command hierarchies in social networks[14].	Bourqui et.al (2009)
3	K-nearest Neighbor	KNN is being used in statistical estimation and pattern recognition since the early 1970's as a non-parametric technique (Saedsayad.com). This technique has not been very popular regarding the sentiment analysis in the social media[12].	Adedoyin-Olowe, M., Gaber, M. M., & Stahl, F (2013).
4	PageRank	TwitterRank as an extension of PageRank algorithm, is proposed to measure the influence of users in Twitter. The experimental results show that the proposed Twitter Rank outperforms other related algorithms. This tries to fix the disadvantages of in-degree and PageRank by considering ink structure and topical similarity among twitter[15].	J. Weng ,E. Lim ,J. Jiang ,Q. He (2010)



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Issue 4, April 2017

5	<b>Markov models</b>	<p>Social Network Discovery based on Sensitivity Analysis [16] introduces an algorithm that computes the Markov property of nodes and the sensitivity parameter. By encoding the pair of nodes depending on magnitude of sensitivity property to predict an edge between the nodes.</p> <p>The algorithm has been tested in the VAST challenge social network data set and the MIT Reality data set. The results proves that the algorithm can be used along with an visualization technique can be used to discover patterns in social networks.</p>	T. Crnovrsanin, C.D. Correa, K.L.Ma (2009).
6	<b>Clustering</b>	<p>Different clustering algorithms including K-Means, Make Density Based, Farthest First, EM and Filtered on Face book 100 university dataset. According to the results of their experiment average accuracy rates for 10 university set are as follows: K-means (64%), Make Density based(55%), Farthest First(60%), EM(60%) and Filtered (57%) [47].</p>	P. Nancy, R. G. Ramani (2012).
7	<b>Classification rules</b>	<p>C&amp;RT algorithm to improve market response by predicating target groups. This has used on real data from the Biznes.net social network .When analyzing the results of the algorithm it proves that applying simple data mining can be used to improve marketing responses</p>	J. Surma , A. Furmanek (2010).

## IV. CONCLUSION

The research scope and the results of the survey papers make some useful sense for controlling the attributes which mostly affect social media. The Data mining techniques provides a better data control facility. The data mining g techniques supports for discovering the similarities among the patterns which exist among the voluminous data set. From the outcomes and the results produced by the researchers makes a new dimension for the researcher to control the uncontrollable data exist in the social medias and social networks.

## REFERENCES

- [1] Clemons, E.K., "The Future of Advertising and the Value of Social Network Websites: Some Preliminary Examinations", Minneapolis, Minneosta, USA, 2007.
- [2] "Social Network Marketing: The Basics", [http://www.labroots.com/Social\\_Networking\\_the\\_Basics.pdf](http://www.labroots.com/Social_Networking_the_Basics.pdf), Aug 01, 2013.
- [3] Raju, E., and Sravanthi, K., "Analysis of Social Networks Using the Techniques of Web Mining", International Journal of Advanced Research in Computer Science and Software Engineering, 2012.
- [4] Rahman, M. M., "Mining Social Data to Extract Intellectual Knowledge", arXiv preprint arXiv:1209.5345, 2012.
- [5] Mosley JR, R. C., "Social media analytics: Data mining applied to insurance Twitter posts", In Casualty Actuarial Society E-Forum, Vol 2 p. 1, 2012.
- [6] Adedoyin-Olowe, M, Gaber, M. M., and Stahl, F., "A Survey of Data Mining Techniques for Social Media Analysis".
- [7] Heatherly, R., Kantarcioglu, M, Thuraisingham, B. and Lindamood J., "Preventing Private Information Inference Attacks on Social Networks", Tech. Rep.UTDCS-03-09, University of Texas at Dallas, 2009.
- [8] Fernando S.G.S., "Empirical Analysis of Data Mining Techniques for Social Network Websites", COMPUSOFT, An international journal of advanced computer technology, 3 (2), Volume-III, Issue-II, February-2014.
- [9] Mariam Adedoyin-Olowe, Mohamed Medhat Gaber, and Frederic Stahl, "A Survey of Data Mining Techniques for Social Network Analysis".



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Issue 4, April 2017

- [10] Umadevi B., Sundar D., and Alli Dr.P., "An Optimized Approach to Predict the Stock Market Behavior and Investment Decision Making using Benchmark Algorithms for Naive Investors", Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on IEEE Xplore Digital Library, pg1 -5, 2013.
- [11] Umadevi B., Sundar D., and Alli Dr.P., "Novel Framework For The Portfolio Determination Using PSO Adopted Clustering Technique", Journal of Theoretical and Applied Information Technology", Vol. 64 No.1, 2014.
- [12] Adedoyin-Olowe M., Gaber, M. M., & Stahl F., "A survey of data mining techniques for social media analysis", arXiv preprint arXiv: 1312.4617, 2013.
- [13] Yang F., Xu Z., Li S., Li X., "Social network mining based on improved vector space model", Proceeding ICIMCS '10 Proceedings of the Second International conference on Internet Multimedia Computing and Service pp 118-121, 2010.
- [14] Bourqui R., Gilbert F., Simonetto P., Zaidi F., Sharan U., and Jourdan F., " Detecting structural changes and command hierarchies in dynamic social networks", IEEE Advances in Social Network Analysis and Mining, pp 83-89, 2009.
- [15] Weng J., Lim E., Jiang J., He Q., "TwitterRank: finding topic-sensitive influential twitterers", WSDM '10 Proceedings of the third ACM international conference on Web search and data mining , pp 261-270, 2010.
- [16] Crnovrsanin T., Correa C.D., Ma K.L., "Social Network Discovery based on Sensitivity Analysis", IEEE Advances in Social Network Analysis and Mining, pp 107-112., 2009.
- [17] [www.wikipedia.com](http://www.wikipedia.com).

## BIOGRAPHY

**Dr.B.UMADEVI** has received her Doctoral degree in Computer Science from Manonmaniam Sundaranar University, Tirunelveli, India. Currently working as Assistant Professor & Head - P.G and Research Department of Computer Science, Raja Doraisingam Government Arts College, Sivagangai-Tamilnadu, India. She has over 22 years of Teaching Experience and published her research papers in various International, National Journals and Conferences. Her research interests includes Data Mining, Soft Computing and Evolutionary Computing. She got the Best Paper Award for her publication in the IEEE International Conference on Computational Intelligence and Computing Research held on 27<sup>th</sup> Dec 2013 at VICKRAM College of Engineering and Technology.



**P. SURYA** is an M.Phil Research Scholar in PG & Research Department Of Computer Science, Raja Doraisingam Government Arts College, Sivaganga, Tamilnadu, India. Her research interest includes in Data mining and Evolutionary Computing.

