



# Social Big Data And Their Challenges

Nilam Deepak Padwal

Research Scholar, Bharati Vidyapeeth (Deemed to be University) Kolhapur

**Abstract:** Social big data comes from two different domains one is social media and another is big data. Social media is very good platform to connect with each other, both together can effectively target their audience for business purpose. Big data becomes well known buzzwords, because velocity, variety, and volume of big data, security and privacy issues are magnified. Social data comes from many different resources but in this paper we focus on social media data. Big data having large number of issue on web storage and social network. The combination of big data technologies and machine learning algorithm has generated challenges in area of social media and social networks and the challenges mainly focus on data processing, data storage, and data representation using pattern mining, analyzing user behavior and tracking the data. \*V model provides definition related to what is big data based problems, application, software or framework. Social media data application can be machined learning, Natural language processing, opinion mining can be used in many different areas such as cyber security, health issues, hobbies, and e-business. Large amount of heterogeneous data, to process that data and generate knowledgeable information from it is one of the biggest challenges. The conceptual model of social big data shows framework of big data technologies. In this paper we discuss current challenges of social big data such as security, privacy issue, streaming and online algorithms, methods of data fusion and data utilization, job opportunities.

**Keywords:** Big data, social media, machine learning, \*V model, data analysis, social issues, PCDM.

## INTRODUCTION:

In social network users can communicate to each other through the transfer protocol used in network system, big data is one of the modules of it. Big data has becomes on large issue when there is large number of storage on web and social network. The combination of big data technologies and machine learning algorithm has generated challenges in social media and social network. The challenges mainly focusing on data processing, data storage, data representation using pattern mining and analyzing user behavior to tracking the data. The data comes from everywhere so managing such large amount of data and extracting useful knowledge from these data source is one of the challenges. Big data is not only storage or access of data but also it's analyzed the data and exploits their value. It can refer the datasets in terabytes to petabytes in size. Theses size can be extend beyond the ability of data base software model and tools to capture store, manage and analyze them efficiently.

## METHODOLOGY:

Big data can be defined through the 3V model, which was defined in 2001 by laney as "high-volume, high-velocity and high-variety information assets that demand cost effective, innovative forms of processing for enhanced and decision making". More recently in 2013 Gartner[4] updated definition as "Big data high volume, high variety and/or high variety information assets that require new forms of processing to enable enhanced decision making insight discovery and process optimization". According to this definition big data having three features volume, variety, and velocity. So 3V model can be extend to 4V including new V that is value or it can be 5V as V veracity. This definition V model defines what is and what is not.

1. **Volume:** refer as gathering all sources of data, different sources of digital devices. so to gathering ,analyzing and processing these data and generate valuable and knowledgeable data for the people it is one of the challenge .transaction based data stores through the year, unstructured data streaming from social media, increasing amount of sensor and machine to machine data being collected so the excising volume was storage issue .
2. **Velocity:** refers to speed of data to transfer. The data content is constantly changing, to introduce of previous data and different form of the streamed data from multiple sources. The new algorithm and methods are requiring to adequate process and analyze the online data from different sources. Reacting quickly to deal with data generation speed is challenging for more organization.
3. **Variety:** refers as variety of data from various sources such as smart phone, sensors or social data. These data contains text audio, video, data logs etc .these data can be structured or unstructured format. Information created from line of business application.
4. **Value:** refers the process of extraction of valuable information it is also called as big data analytics. it is most important application of big data which generate useful in information. Identification of value from big data application area are still challenge.
5. **Veracity:** refers the correctness and accuracy of information which manages data quality, data governance and metadata management which their legacy and privacy concern.



Big data sources are open science, data cloud, health care data, public dataset on amazon's web services etc. as social media is most representative, relevant data source, which generated data from wide number of internet application. Website such as Facebook, Twitter, LinkedIn, YouTube, Instagram, Google, tumblr flicker etc. these websites allow user to connect, interact and share the collaborate the data such information has spread many different area in everyday. Social big data comes from two different domain i.e. .social media and big data. Social media is basically the large amount of data that comes from different sources but we mainly focus on social media data. its application can be machine learning, natural language processing, opinion mining, can be used in many different areas such as cyber security, health ,hobbies , e-business etc.

The definition of social big data is " those process and method that are designed to provide sensitive and relevant knowledge to any user or company from social media data sources can be characterized by their different formats and contents, their very large size ,and the online or streamed generation of information".

When there is large amount of data heterogeneous data which can be unstructured or semi structured data . To process and analyze such data and generate useful knowledge from it is one of the biggest challenges of big data in social media.

There are some classic methods, algorithm, and framework not able to solve such problem completely. The biggest challenges are knowledge representation, data management, data processing, data analysis, data visualization. These challenges are including,

- 1] **Management issue:** accessing very large quantities of unstructured data.
- 2] **Quality V/S quantity:** how much data is enough for the high quality data?
- 3] **Ownership and security:** processing the dynamic streaming is continuously changing

How to identify valuable data to analyze that data and discover useful knowledge so its improve the decision making for users. The conceptual map of social big data, following figure shows three area of conceptual model.



1. **Social media:** Natural source of data analysis.
2. **Big data:** parallel and massive processing paradigm.
3. **Data analysis:** set of algorithms and methods used to extract and analyze knowledge.

The intersection between big data and data analysis shows machine learning framework of bid data technologies.

The intersection between data analysis and social media is the concept of web based application which uses social media data; Ex: marketing and E-health. The intersection between social media and big data reflects social media applications such as LinkedIn, Face book, YouTube etc. which uses big data technologies like hadoop to develop their web system.

The center of this figure represents goal of social big data i.e. knowledge extraction and exploitation. In this paper we have to discuss the current challenges of social big data.

1. **Privacy issue:** the online big data and social media suffer from privacy issue as protecting user privacy on social media is challenging, now there are many privacy issue.

There are some privacy preserving techniques which addressing privacy related issues with following approaches.

- 1.1. **K-anonymity:** which is a property possessed by certain anonymised data , given the private data and set of specific field. the system has to make the data practically useful without identifying the individual who are subject to the data.
- 1.2. **Differential privacy:** This can provide efficient way to maximize the accuracy of queries from statistical database while minimizing the chance of identifying its records.



**2. Streaming and online algorithms:** To find the adequate approaches to analyze the massive amount of online data or stream data is one of the big challenge in big data problems. Classification require alots of efforts for real time analysis. Classification and clustering are technique the best type of algorithm used to achieve the analysis goal. data stream can be categorized in two ways.

**2.1 Offline analysis:** in this large amount of data and applying clustering algorithm to Analyze the data.

**2.2 Online analysis:** these are real time data, in these algorithm continuously receiving the data can not able to keep pass information To manage the social big data challenges require high scalability in terms of memory consumption some new modification with traditional Clustering algorithms, such that K-Means, EM so that they work with Map reduce paradigm.

**3 Methods of data fusion and data visualization:** These are clear challenges in big data.

**3.1** In data fusion require some reliable methods to access multimedia objects for social media application.

**3.2** To study the dynamics of individual group behavior characteristics, pattern of information diffusion, identifying influential individuals in social network and other social based applications.

**3.3** To produce high quality clustering results, identifying events in social media via clustering and similar metrics learning approaches.

**3.4** The capacities to collect and store new data, to analyze this data volume, record the data about movement of peoples and objects at large scale. These are the challenges related visual analytics.

### CONCLUSION:

Social big data comes from different available resources and secure communication with this data suffer from social identification and evaluation with such data and privacy preserving services with data is not easy at all .so in future we have to set some useful benchmark dataset for evolution. To change the data in such way that changes made by data mining algorithm should not compromised security of data is called as privacy –concern data mining (PCDM) .data Security of big data is gaining momentum in research community is due to emerging technology like social network. Data privacy suffer in many other research area such as context of law , enforcement ,illegal usage etc. the social impact of the big data is direct connection between capacity of big data application to promote the bold action and solution to societal challenges or problems. In this aim of study investigate the social challenges are security, socially inclusive economic growth ,access education or quality of education ,equal opportunities for all, job opportunities. When we are talking about social big data its very difficult as in future as trending topic in future.

### ACKNOWLEDGMENTS:

I would like to thank Bharati Vidyapeeth,(Deemed to be University) research reported through the center Kolhapur.

### REFERENCES:

1. Social big data: Recent -achievements and new challenges by Gema Bello-Orgaza , Jason J. Jungb ,David Camachoa , Elsevier- 17 august 2015.
2. Big data Analytics: Challenges and Applications for Social Media Data by Krishnaveni , Udhayakumar, NCETCT-18,2018
3. Big Data and Their Social Impact: Preliminary Study by Miltiades D. Lytras and Anna Visvizi. MDPI 17 September 2019.
4. M. A. Beyer, D .Laney, The Importance of 'Big Data': A Definition, Gartner, Stam-ford. Published: 21 June 2012, ID:G00235055.
5. Privacy and Security of Big Data: Current Challenges and Future Research Perspectives by Alfredo Cuzzocrea. CIKM: 07 November 2014.
6. Towards the Development of Best Data Security for Big Data by Yuan Tian , CN Vol-9,No .4 November 2017.
7. Social issues of big data and cloud privacy confidentiality and public utility by Koichiro Hayashi.ijcaonline.org>NACA-2015.
8. [www.upgrad.com/blog/major-challenges-of-big-data/19/05/2020](http://www.upgrad.com/blog/major-challenges-of-big-data/19/05/2020).
9. [www.scnsoft.com/blog/big-data-challenges-and-their-solutions/21/03/5018](http://www.scnsoft.com/blog/big-data-challenges-and-their-solutions/21/03/5018).
10. <https://www.scnsoft.com/blog/big-data-security-challenges04/03/2018>.
11. Social Networking Data Analysis Tools & Challenges by Androniki Sapountzi and Kostas E. Psannis. Future Generation Computer Systems, 2016, ISSN: 0167-739X.
12. <https://insidebigdata.com/2018/10/06/4-major-ways-big-data-impacting-social-media-marketing>.
13. [https://www.researchgate.net/publication/306034329\\_impact\\_of\\_big\\_data\\_and\\_social\\_media\\_on\\_society/link/57abec7308ae7a6420bfe679/download/11/08/2016](https://www.researchgate.net/publication/306034329_impact_of_big_data_and_social_media_on_society/link/57abec7308ae7a6420bfe679/download/11/08/2016) .
14. Impact of Social Media Networks Big Data Analysis for High-Level Business T. Veeramani, P. Srinivasarao, B. Rama Krishna, R. Thilagavath. January 2019.
15. D.Laney,3D Data Management: Controlling Data Volume, Velocity, and Variety, Technical Report, 2001. URL <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (accessed August 2015).
16. L. Sweeney, K-anonymity: a model for protecting privacy, Int J. Uncertain. Fuzziness Knowledge-based Syst. 10(5)(2002)557–570.
17. C. Dwork, Differential privacy: a survey of results, in :M. Agrawal, D. Du, Z. Duan, A. Li(Eds.) ,Proceedings of 5<sup>th</sup> International Conference on Theory and Applications of Models of Computation (TAMC2008), Xi' an, China, April2529, Lecture Notesin Computer Science, 4978, Springer,2008,pp.1–19.
18. W. Y .Chen, Y. Song, H. Bai , C. J. Lin, E .Y. Chang, Parallel spectral clustering distributed systems, IEEE Trans.PatternAnal.Mach.Intell.33(3)(2011)568–586.
19. .D. Menendez, D. Camacho, Gany : a genetic spectral-based clustering algorithm for large data analysis, in :IEEE Congress on Evolutionary Computation (CEC), 2015, IEEE, 2015, pp.640–647.