Researchpedia
Open Access Journals

Received: 27 March 2020; **Revised:** 04 April 2020; **Accepted:** 04 April 2020; **Published:** 10 June 2020
Researchpedia Journal of Computing, Volume 1, Issue 1, Article 5, Pages 39-49, June 2020

Digital Object Identifier RpJCS.2020.101.DOINumber

A Review on Big Data and Social Network Analytics Techniques

Muhammad Nouman Noor[1], and Farah Haneef[2]

[1]Department of Computer Science, Hitec University, Taxila, Pakistan
[2]Department of Computer Science, Quaid-e-Azam University, Islamabad, Pakistan

Corresponding author: Muhammad Nouman Noor (e-mail: mnoumannoor93@gmail.com).

ABSTRACT Social media and social networks assist people to furnish opinion and to provide feedback about a product or a service, which pay immense influence on the businesses. By utilizing such data, businesses may grow up or mislay their customers and revenue generation. The data generated in a social network may vary enormous and may not be analyzed by traditional approaches. In this regard, machine learning techniques help businesses to be mature. Machine learning is a domain which provides schemes to learn and extract important patterns from the data, without being program explicitly. This paper provides opportunity for upcoming researchers to find research gaps about machine learning techniques those have not applied so far for social network analytics and also provided discussion / summarization of tools that are effective for Social Network Analytics and Big Data.

Keywords Big data, data analysis, machine learning, predictions, social networks.

I. INTRODUCTION

The term Social Network is almost known to every individual or at least every IT professional. This term best described as an online service that allows creating individual profile in order to communicate and share videos, pictures, knowledge and interests with the other users in the network like Facebook, Twitter and Linked-in [1]. The fact that social networks gain huge success because they provide opportunities for individuals to connect people having similar interests, share knowledge and do learning [2]. Social networks are vital sources for online interactions, sharing of contents [5], subjectivity [6], assessment [7] stand out in blogs, text, discussions, reviews, remarks, news, reactions or some other documents [8]. Social networks provide platform to users for rapid exchange of information regardless of their location [1]. Social network represents social relations in the form of graph; links and nodes [3]. Entities are represented in the form of nodes and the relationships among these entities form links [4]. Networking websites also provide a platform for celebrities, private organizations, senior government officials and government organizations to get knowledge about the feedback of their audience regarding uploaded posts therefore; they are concerned with massive data generated from social networks [1] as shown in fig 1.

The beginning of big data analytics brings a revolution in businesses. It has changed the way, people and businesses look at the data [9, 10]. Big data has established a competitive advantage among rivalry organizations in a way they utilize big data for important business decisions [11].

A very common misconception existed in the society that any large amount of the data can be seen as big data. In reality, data generated by any method which satisfies the restrictions imposed in the dimension of *velocity*, *volume*, *variety* and *veracity* can be called as Big Data [12] [69] as shown in fig 2. Data which is produced at a larger rate have many different attributes can be utilized for learning and making decisions then it falls under the umbrella of big data. The big data processing is different in a way conventional data is handled. Processing of Big data requires special techniques/tools to make an efficient analytical decision [11]. In the current era, as the internet is available around the globe and the world is moving towards IOT, big data existence shows in almost every aspect of life. This big data is used by data scientists/organizations to make important decisions such as business processes optimization, target consumers and performance appraisal in different domains; hospitality industry, health care, learning systems and education, agriculture, automobile [13]. In order to gain further insights to the big data applications, this paper describes the machine learning techniques for social network analytics.

Machine learning is evolved from theory of computational learning in Artificial Intelligence [18]. Machine learning provides system with ability to learn automatically through data and extract important information without being programmed explicitly [14]. The extraction of important information is based on finding the patterns in the data and make

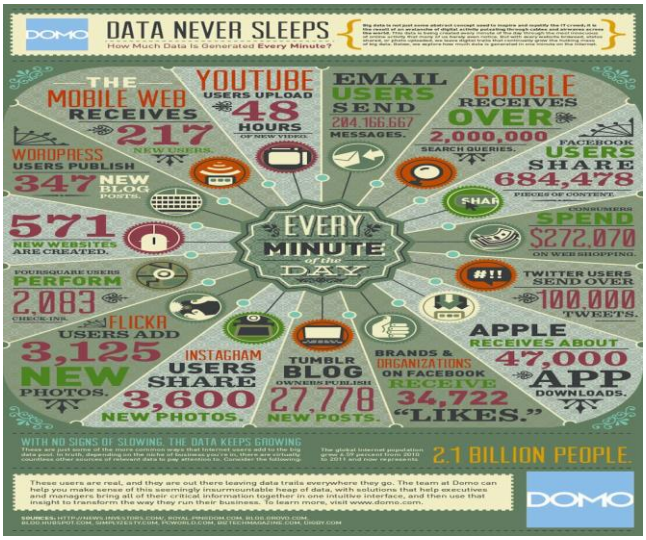

FIGURE 1. Estimated Data Generates on Social Networks Every Minute [1].

predictions accordingly. In today's world, techniques of machine learning have been used in various complex fields such as biology, medicines, Social Media and astronomy in order to find the hidden information inside the data [16]. Learning from the various data is likely to bring major progress in engineering and sciences in order to improve our quality of life [17]. Machine can be learned in different settings; Reinforcement, Unsupervised and Supervised [15]. *Reinforcement learning* is a domain in which learning is done by interacting with system, taking different actions and getting their results [19]. In *Unsupervised learning* learner learn from input data that is unlabeled [20]. *Supervised learning* is a domain in which learner use labeled input data for learning [21].

The purpose of this paper is to explore the techniques used to analyze the immense data for efficient predictions. Related studies [68][69][70] mostly dealt with specialized attributes; event detection techniques, finding influential bloggers or deep learning techniques for social media analytics. The author of [25] used the techniques of social network analytics for Churn prediction in Telecom Industry. In this paper, we discussed general techniques of machine learning for social media analytics and big data along with those techniques which are not considered till now.

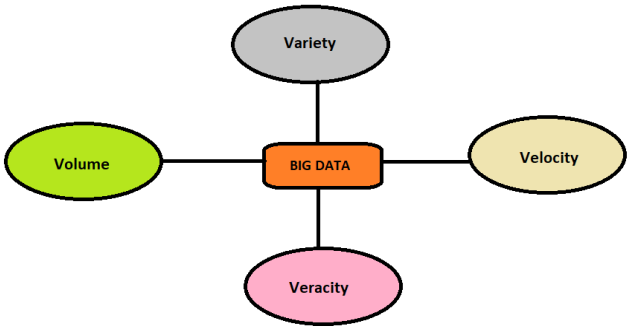

FIGURE 2. Big Data Attributes [11].

The remaining sections of this paper are organized as follows; section 2 describes concepts and terminologies related to big data, machine learning and social network analytics. Section 3 discussed the importance of social network analytics. Section 4 elaborates the brief analysis of general data. Section 5 describes big data tools and techniques along with social network analytics. Section 6 illustrates machine learning techniques for social network analytics. Section 7 discussed the tools for Social network analytics. Section 8 briefly describes dataset and at last we conclude the paper in section 5.

The main contributions of this paper are given below

- Discussion of techniques and practices of machine learning, big data and social network analytics.
- Exploration of machine learning classifiers those can be applied for social network analytics.
- Discussion and Summarization of big data tools.
- Discussion and summarization of built-in and market available social network analytics tools.

A. METHODS FOR SELECTING PAPERS

1) SELECTION STRATEGY

Keeping the research objectives in our mind we have selected the relevant papers on the basis of following criteria.

- We have included articles those are relevant to machine learning techniques for social network analytics.
- We have incorporated articles those are related to big data analytics techniques.
- Those articles in which importance of social network analytics is discussed are also incorporated.
- We have also incorporated articles those are related to big data and social network analytics tools.

2) QUALITY CRITERIA

To ensure the quality of included papers, we have considered only those models/techniques which describe their data set and/or evaluation criteria. We ensured that these papers provide statistics to support their analysis and evaluation.

3) DATA SOURCES

We acquired appropriate papers from the following data sources:

- ACM Digital Library
- IEEE Xplore
- Springer Link
- Science Direct
- Google Scholar
- Elsevier
- Online Resources

4) SEARCH QUERIES

To meet our research objective, we have used search queries for each data source by considering prime concepts and their synonyms. Those search queries were: Big data techniques for data analysis, techniques for data prediction in machine learning domain, machine learning techniques for data analysis, machine learning techniques for social network analytics, importance of social network analytics, big data tools available, and social network analytics tools.

II. CONCEPTS AND TERMINOLOGIES

In current section, we will provide explanations of concepts and terminologies of Machine Learning.

As statistical techniques lack the prediction power, therefore we require Machine learning techniques for learning and predicting unseen data. Three types of learning are:

A. SOCIAL NETWORK ANALYTICS

Users of social media share immense amount of information and general content on social media platforms like Facebook, Twitter and Linked-in as images, text, tags and videos. Social network analytics is the analysis of such content. This information can be useful for some businesses or getting feedback of certain communities or general public on certain product, service or slogans.

B. BIG DATA

A large amount of data which can be examined using computations to find the trends or patterns within data especially that is associated with public behavior as in social media. Data which fulfills the limitations imposed in the dimension of *velocity*, *volume*, *variety* and *veracity* can be called as Big Data.

1) VOLUME

Volume can be termed as the size/quantity of data which is generated, collected and stored. Data quantity tells us whether the data can be called as Big Data or not.

2) VARIETY

Variety means type or form of data. Big Data can be classified as Images, Audio, Text or Video.

3) VELOCITY

Velocity means speed at which data is generated / analyzed. Big Data is mostly generated in a real time like the data in social media which changes every second.

4) VERACITY

Veracity is referred as the quality of data. Quality of data determines the accuracy of analysis.

C. MACHINE LEARNING

Machine learning is the branch of artificial intelligence in which machine or system can learn from data, find its patterns and make decisions on the basis of this learning.

D. UNLABELED DATA
Unlabeled data consist of samples with no meaningful tags, labels or class associated with it that can describe the data. For example, horse image does not have caption which machine can understand like "horse image". Unlabeled data might include photos, videos, and audios with no caption or explanation.

E. LABELED DATA
Labeled data is a collection of unlabeled data with some meaningful caption or information in the form of "Class", "Tag" or "Label" associated with its all instances. For example, image of horse or buffalo has caption like "Horse Image" or "Buffalo Image".

F. SUPERVISED LEARNING
Supervised Learning is machine learning method which is done on training data which has classes or labels associated with it and produce objective function which classifies other unseen data [68].

G. UNSUPERVISED LEARNING
Unsupervised Learning is a machine learning method which is done on training data which don't have classes or labels associated with it and forms cluster of data on the basis of similarity [68].

H. REINFORCEMENT LEARNING
Reinforcement Learning is a machine learning technique which is done by taking action and recording its feedback and then carrying out learning on this data [54].

I. CLASSIFICATION
This method classifies the data into different classes. Just as the number can be zero or one not both [0,1].

J. LINEAR REGRESSION
This method has continuous outcome means it just give us the exact number.

K. LOGISTIC REGRESSION
This method has discrete outcome and it classify the machine learning data. It gives us probability between zero and one [0-1].

III. IMPORTANCE OF SOCIAL NETWORK ANAYLTICS
Social network analytics is useful for almost every field for making decisions on the basis of social trend. Businesses need social network data analytics for tracking their product reviews, politicians for finding public reputation, governments for getting public opinion about certain campaigns and health sector for checking medicines popularity and its effectiveness. Some of the main benefits of Social Network Analytics are given below.

- It helps in providing extra customer care by evaluating the needs and interests of customer on social networks.
- Enable us to identify target audience, target community or hidden customers.
- Helps in analyzing the response of general public on steps that are taken during the time of crisis.
- Response of general public for new ideas, products to pursue them.
- During the days of elections, helps in predicting the voters' intentions.
- Tracking of trends in specific community.
- Monitor the health of brand/business.
- Identify the influencers of community to get partnered with.
- Save time on taking critical decisions.
- Enable businesses to get ahead with their competitors by comparing the products/services.
- Helps in improving Return on Investment (ROI).
- Detect opponents' challenges and gain their unhappy customers.
- Helps in generating effective strategy.
- Getting and evaluating timely feedback.
- Predicting future of certain product or service.
- Finding popularity of certain personality.
- Determine whether the product or service gaining popularity or declining with time.
- Fraud detection of companies working in health insurance sector.
- Helps in analyzing Crime patterns.
- Helps in counter terrorism.

- Identifying the religion extremists.
- Financial corruption investigation.
- Recommending the interested content/product to the right audience.
- Provide guidance about the career or subjects to students.

IV. DATA ANALYSIS TECHNIQUES

As per Technorati, approximately 1.2 million posts and 75,000 blogs are generated every day that are providing feedbacks and giving suggestions about products & services [22]. The huge amount of data that is generating on social networks every minute makes it difficult to exercise traditional approaches like statistical techniques for the purpose of analysis [23]. A writer in [24] proposed 5 different techniques to predict event for 5 different datasets; opinion based, volume based, distribution of events, tracking of activities & cause of protest. In paper [25] the author discussed many common techniques that are used to detect content of hate promoting i.e. Naive Bayes, Text Classification, KNN, Rule Based Classifier, SVM, Decision Tree, Exploratory Data Analysis (EDA), Clustering (Blog Spider) and Keyword Based Flagging (KBF). The author also discussed techniques to locate and identify users with similar interest including Link Analysis and Topical Crawler. The author of paper [26] provide us with a techniques i.e n-gram, Language modeling and boosting to mine text data and classifying hate speech content. In the paper [3] different data mining techniques; supervised learning, semi-supervised and unsupervised methods are used to analyze social networks. In [27], author proposed a model named as latent variable model for estimation of relationship strength in social networks. The experiment done by the author shows that graphs created by the 'estimated relationship strength' will give better classification performance and higher auto-correlation than the graphs developed by numerous characteristics of raw data.

Singla [28] examine the correlation between individual seek matters among people that interact with instant messaging and demonstrate that correlation exist which improves with the number of times they communicate. The novel model called "multi-feature space latent Dirichlet allocation (MFS-LDA)" is discussed for recommendation of tags [71]. The author of [73] studied the relation among subjects and professions in a network science perspective which is also an emerged field. As we see many methods are already being used for opinion and sentiment analysis, categorizing on the basis of opinion text with the use of binary division (Positive against negative) [29], [30], [31], [32] is found insufficient when recommend items.

V. BIG DATA TOOLS, TECHNIQUES AND SOCIAL NETWORK ANALYTICS

Social Network Analytics is recently included in diverse areas i.e. Medical, Security and Telecom. It could be used for detecting communities of consumer, coaches who can attract others to acquire customers in the social network [67].

Social Networks including Twitter, Facebook and linked-in generate huge amount of data every minute, the data requires analytics to obtain information which may fruitful for businesses. It mainly helps to solve the newly created problems or any old problem in better/effective way [45], [46], [47], [48]. The scrutiny of data both structured and unstructured paved the way towards social network analytics [33]. Big data analytics techniques include all of the following; predictive analytics, statistical analysis, data mining, data visualization, complex SQL, natural language processing and artificial intelligence [11]. In this section, we will discuss various big data tools and social network analytic techniques for processing user generated data on social networks.

Various tools which are used for big data analytics are, i) Apache Hadoop [34], an open source tool which is reliable and powerful [35], consist of components known as Map Reduced Framework & Hadoop Distrbuted File System [36]. ii) Apache Storm [37] & iii) Apache Stark [38] mainly used for real time distributed computation [35]. iv) Apache Hive, exercise for analysis, query and data summarization [39]. v) Jaql, utilize for functional data processing. vi) NoSQL. Hama and Spark tools are also gaining popularity in social network analytics research [50]. Big Data be able to analyze via text analytics, predictive analytics, data mining and statistical analytics [40], [41], [42].

Big Data Analytics types as discussed in [43] are i) Prescriptive, it helps to choose which action be supposed to take ii) Predictive, helps in predicting future iii) Diagnostic, analyze the past situation, why it happened and how to overcome iv) Descriptive, analyze current situation and predicting near future. By the help of suitable analytic technique, businesses can predict their future in order to gain more profit or improve their services. Predictive analytics helps businesses to make faster and better decisions [44]. A framework proposed by the authors of [72] called as "CommuniMents", can actually calculate the sentiments of a public about a certain occasion. CommuniMents used "automated snowball sampling" which find the public participants, and after that gets the content which participants posted about certain occasion. Then it processes their posts/tweets and calculates the sentiments of the public. A study was conducted by the author of [74] about the requests of blood donation on social media. They have analyzed almost 900 accounts that have following of more than 35,000 users.

The author of [11] guides us that Social networks can be analyzed by two approaches i.e. graph database approach and parallelization approach [49]. The main center of attention in parallelization approach is to divide massive dataset into small subsets and compute the data in concurrent manner via cloud computing. Graph databases mostly use for networks where the data structure directly represents the key characteristic of the problem. Adjacent elements in these databases are interconnected. Link among nodes is extremely helpful while dealing with vastly interlinked social network data [51]. Their Scalability to big datasets is high because they don't need costly join operations [52]. Due to these features, graph databases became an excellent

option when working with social network analytics. Broad collection of graph databases that are used in social networks analysis are ArrangoDB, Allegrograph, NoSQL, MongoDB, Cassandra, Neo4j [53].

In Table 1, various tools and techniques of Big Data are summarized and presented in compact form.

TABLE I. Summary of Big Data Tools and Techniques

Tools	Description
Apache Hadoop	An open source tool which is reliable and powerful, consist of components known as Map Reduced Framework & Hadoop Distributed File System.
Apache Storm & Apache Stark	Mainly used for real time distributed computation.
Apache Hive	Exercise for analysis, query and data summarization.
Jaql	Utilize for functional data processing.
NoSQL	Mainly used for storage/retrieval of data. In NoSql, data is displayed/ showed not in the form of tabular relations as shown in relational databases.
Hama and Spark tools	Recently gaining popularity in social network analytics research.
Techniques	**Description**
Prescriptive	It helps to choose which action is supposed to take.
Predictive	Helps in predicting future.
Diagnostic	Analyze the past situation, why it happened and how to overcome.
Descriptive	Analyze current situation and predicting near future.

VI. MACHINE LEARNING TECHNIQUES FOR SOCIAL NETWORK ANALYTICS

As discussed in the last section about the big data applications for social network analytics, this section aims to discuss the techniques of machine learning that can apply to social networks data.

The online opinion of any person can influence the public image of brand and its profitability therefore the businesses are now more concern about what are people saying about their product or services on social networks. This fact has raised the field of social network analytics [55]. When social networks data is exercised for such an important decision making, it turns out to be necessary to process the data that is obtained from online Social Medias by the help of machine learning techniques.

Machine learning is the process to provide machine the ability to learn by feeding the data as an input to it. In [56] the author discussed that the arrival of big data has improved machine learning techniques so much that complex calculations can be applied to massive data. Machine can learn by the help of three techniques that are i) Supervised Learning, the training data is labeled ii) Unsupervised Learning, the training data is unlabeled iii) Reinforcement Learning, perform action and get feedback [54]. Social Networks generally analyzed by unsupervised learning as the data of social media is unstructured [57]. Machine learning techniques are mostly applied on any system in order to make predictions. Machine learning techniques can be classified in different types that are i) Classification, ii) Linear Regression, iii) Logistic Regression iv) Neural Networks. Machine learning techniques/ algorithms that plays a main role in social network analytics include Naive Bayes, Decision tree learning, Maximum Entropy method, Nearest Neighbor classifier, Dynamic Language Model classifier, Support vector machine, Simple logistic classifier, linear regression & logistic regression, Multilayer Perceptions and Bayes Net [11].

Naïve Bayes is a classifier mostly used for classification of text in order to find the category of text/document i.e. sexual content detection, spam emails detection [58]. In various research papers [59] [60] it is shown that when Naïve Bayes, Support Vector Machines and Maximum Entropy Method used for supervised learning, Naïve Bayes performs well but when the same used for unsupervised learning, Support Vector Machine outperform Naïve Bayes. Decision tree also used for classification of text; it has root node that have all the documents. The algorithms [63] that search through these trees are ID3 [61] and C4.5 [62]. Nearest Neighbor classifier is a supervised learning method for pattern recognition with in data. This technique is utilized in social network analytics when there is no or less information about the shared data. Support Vector Machine (SVM) is a learning algorithm that exercises on supervised learning (Labeled Data) and analyze data which is ultimately used for regression analysis and classification [64]. Language model classifier is exercise to classify test data into two classes. Language model classifier appears as an extension of Naïve Bayes since both of them uses the identical hypothetical framework [65]. Linear and Logistics regressions are used for predictions. Linear regression finds the one single output value but the logistic regression uses probability to show the output value. In the beginning of social network analytics field, logistic regression and logistic models applied for analysis which was based on

graph theory [66]. Multilayer Perceptron works just like simple Perceptron to classify data but have multiple layers, in which every previous layer is completely connected with its next layer.

In Table 2, various techniques of Machine Learning which are used for social network analytics are summarized and presented in compact form.

TABLE II. Summary of Machine Learning Techniques Used for Social Network Analytics

Techniques	Method	Purpose of Use
Naïve Bayes	Supervised and Unsupervised Learning	Classification of text to find the category of text/document i.e. sexual content detection, spam emails detection.
Decision tree	Supervised and Unsupervised Learning	Mostly used for classification of Text.
Nearest Neighbor classifier	Supervised Learning	This classifier is a method for pattern recognition with in data. This technique is utilized in social network analytics when there is no or less information about the shared data.
Support Vector Machine (SVM)	Supervised Learning	Support Vector Machine (SVM) is a learning algorithm that exercises on supervised learning (Labeled Data) and analyze data which is ultimately used for regression analysis and classification.
Language Model Classifier	Supervised and Unsupervised Learning	Language model classifier is exercise to classify test data into two classes. Language model classifier appears as an extension of Naïve Bayes since both of them uses the identical hypothetical framework.
Linear and Logistics Regressions	Supervised Learning	Linear and Logistics regressions are used for predictions. Linear regression finds the one single output value but the logistic regression uses probability to show the output value. In the beginning of social network analytics field, logistic regression and logistic models applied for analysis which was based on graph theory.
Multilayer Perceptron	Supervised Learning	Multilayer Perceptron works just like simple Perceptron to classify data but have multiple layers.

Vast variety of machine learning techniques have been applied in the context of social networks but still there are many models which have not been used for social networks analytics so this review provides researchers with an opportunity to find research gaps and implement them.

VII. SOCIAL NETWORK ANALYTICS TOOLS

As we discussed in section-v that Social networks/media can be analyzed by two approaches i.e graph database approach and parallelization approach. In this section we will discuss the various tools available for social network analytics. Choosing the right analytics tool is very important for businesses and mostly depend on the basis of Ease of Use, Supported Social Networks, Supported Data, Speed, accuracy and how often company updates its software. First we will discuss the social network analytics tools which are provided by social platform themselves and then market available tools.

A. FACEBOOK INSIGHTS

This platform provided by the Facebook can be used to see the insights and reachability of posts that are unpaid and paid. It also helps in reviewing the demographic information about our target audience. This tool doesn't provide the sentiment analysis.

B. PINTEREST ANALYTICS

This tool provided by the Pinterest. It helps in measuring the website traffic and also communicates insights about the content that people saved from our profile. This platform provides the demographic insights about the audience and their interests. We can also compare our audience with general users of the Pinterest using this tool. This tool also doesn't provide sentiment analysis.

C. TWITTER ANALYTICS

This platform is provided by the Twitter which helps users get knowledge about the tweets or trends whether these are popular in public or not. This tool also gives you an idea about the demographic information of target audiences which support or oppose the tweet. Additional Analytics which is provided by the platform are Conversation tracking and insights of video viewers. This tool doesn't provide messaging facility to target audience.

D. LINKEDIN ANALYTICS

LinkedIn Analytics provides us with opportunity to see demographics information of audience regarding job functions. It also shows the updates regarding visitors, posts and followers.

E. INSTAGRAM INSIGHTS

Instagram Insights is the most versatile built in social network analytics tool that provides us with social network analytics of data which is growing every second like Posts, Recent Activities, Stories, Profile Visits or Website Clicks. This tool can also

be integrated with Pinterest platform for better social network analytics. It also gives us the Demographics information of our followers [76].

It is important to note here that every social network platform doesn't provide us with built-in analytics tool and mostly built-in analytics tool don't provide us with sentiment analysis. What if our intended audience is present on the social network which doesn't provide us with built-in analytics tool? For this purpose, a lot of market available tools are present to help in this situation.

F. BOARDREADER

BoardReader is a social network analytics tool which helps us for analyzing the forum posts and reviews, an outdated form of social networks. This tool provides with the graphs with which we can compare the conversation volume of our product against competitor's products. But it doesn't provide us with the ability to check whether the conversation about the product is good or bad. It is good for small businesses to get an idea about social network analytics tools. This tool can also integrate with other tools for advance analytics.

G. SQUARELOVIN

SquareLovin is a social network analytics tool which is Instagram focused. This tool offers to analyze growth rate of our posts, monthly report on analysis and posts history by hour, day, month, and year. It also offers to view insights about our target community inclinations and area of interests. This tool helps us for finding the time in which the posts are best boosted and worst neglected.

H. SUMALL

This tool is a versatile and allows us to connect different social platforms like Twitter, Instagram and Facebook. It also helps us to automatically engage with our audience. It offers both in-depth and in-breadth social analytics services.

I. UNIONMETRICS

This tool can provide the analytics of Instgram, Facebook and Twitter. This tool allows us to monitor all subjects, issues and social accounts that are in our interests and helps in developing strategies for social media or accounts.

J. VIZIA

This tool is used to analyze video content from social networks like YouTube. It allows us to increase the views of our videos and get our content viral. This tool would work perfect by integrating with other social platforms that allows sentiment analysis to promote or recommend videos to the right audience. This tool helps in improving our content by providing us with the ability to take feedback from audience in the form questionnaires or polls.

K. WISELYTICS

This tool till now can provide only Facebook analytics. By using this tool, we can track that which posts reach to the most audience and can track the actions, interactions of audience about the posts. It also helps in analyzing about the community that gives negative feedbacks. In Table III, various built-in and market available tools which are used for social network analytics are summarized and presented in compact form.

TABLE III. Summary of Tools for Social Network Analytics

Tool	Supported Platform	Limitations (Tool doesn't Provide)	Availability
Facebook Insights	Facebook	Sentiment Analytics	Built-in on Facebook
Pinterest Analytics	Pinterest	Sentiment Analytics	Built-in on Pinterest
Twitter Analytics	Twitter	Messaging Facility for Target Audience	Built-in on Twitter
Linked-in Analytics	Linked-in	Sentiment Analytics	Built-in on Linked-in
Instagram Insights	Instagram, Pinterest	Sentiment Analytics	Built-in on Instagram
BoardReader	Facebook, Instagram, Linked-In, Pinterest,	Whether Users Comments are Positive or Negative	Market-Available

	Twitter		
Squarelovin	Instagram	First month analytics not available	Market-Available
Sumall	Twitter, Instagram, Facebook	Detail analytics are just for Twitter	Market-Available
UnionMetrics	Twitter, Instagram, Facebook	Multiple accounts difficult to manage	Market-Available
Vizia	Youtube	Some feedback responses discarded automatically	Market-Available
Wiselytics	Facebook, Twitter	Campaign Segmentation	Market-Available

VIII. DATASET

In this study, we have seen that some studies are based on interpretation and have no evident support of data therefore we present them in descriptive form. Mostly authors used the data of twitters for getting tweets. Some authors use Facebook and Instagram data for getting tags, posts, comments and likes. Authors have also used the published datasets of some researchers. There are many APIs available for collection of data from Facebook, Twitter, Linked-in or any other social network like Twitter API and Facebook API. Linked-in provides two types of APIs for data collection; JavaScript and REST API. Authors also used paid datasets like Gardenhose and Firehose for getting twitter data [68]. After using these datasets, Machine learning techniques like Naïve Bays, Decision Trees and Support Vector Machine are discussed and used by different authors for analyzing the social network analytics.

IX. RESEARCH CHALLENGES

As social network analytics is inter-disciplinary field, so many researchers of different fields use it for different purposes therefore; authors discussed and provided different research challenges related to their sub-field on analysis of social networks. Some authors provided challenges related to collection of data and some authors have provided light towards the challenges of analysis. But analysis showed that nearly every author mentioned “Data Volume” as the challenge in their research on social network analytics. It is pertinent to mention here that the real time nature of data is an important aspect while working with social networks as the chances of noisy and inconsistent data became high which can be categorize as important challenge. The data of social network has unstructured and uncertain nature which gave lights to new challenges like quality and structure of data [77].

X. CONCLUSION

This paper provides review of the machine learning, big data and social network analytics techniques. Tools for Big Data and Social network analytics were also discussed and summarized. As per the review, there are some machine learning techniques those are not discussed and used for social network analytics which may interpret the data more accurately for example PART. According to best of our knowledge PART classifier is not use for social network analytics. Therefore, researchers should accommodate this for research. Further work is required to identify the machine learning techniques those needed modifications for social network analytics which may combine with data mining techniques to perform better in analyzing social media.

REFERENCES

[1] Mariam Adedoyin-Olowe, Mohamed Medhat Gaber and Frederic Stahl, “A survey of data mining techniques for social network analysis”, Journal of Data Mining & Digital Humanities, 2014 (June 24, 2014).

[2] Md.Riyazuddin, Shaik Rasool, Syed Azar Ali, Khaja Zahooruddin Ahmed, “Privacy preserving data mining analysis in online social networks (OSNs)”, International Journal of Computer Science & Communication Networks,Vol 5(6),379-385.

[3] D. Kavitha “Survey Of data mining techniques for social networking websites”, International Journal of Computer Science and Mobile Computing IJCSMC, Vol. 6, Issue 4, April 2017, pg.418 – 426.

[4] Asur, S., and Huberman, B., "Predicting the future with social network." Web Intelligence Agent Technology (WIIAT), 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010.

[5] Thompson, J B, “Media and modernity: A social theory of the media”, John Wiley & Sons, 2013.

[6] Asur, S., and Huberman, B, "Predicting the future with social network." Web Intelligence and Intelligent Agent Technology (WIIAT) 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010.

[7] Kim, Y., Hsu, S-H., de Zuniga, H.G., “Influence of social network use on discussion network heterogeneity and civic engagement: The moderating role of personality traits”, Journal of Communication 63.3, 498-516, 2013.

[8] Bakshy, E., Hofman, J. M., Mason, W. A., Watts, D. J., "Identifying influencers on twitter", In Fourth ACM International Conference on Web Search and Data Mining (WSDM), 2011.

[9] "Advancing Discovery in Science and Engineering", Computing Community Consortium. Spring 2011.

[10] Will Dobbie, Roland G. Fryer, Jr., "Getting beneath the veil of effective schools: Evidence from new york city", NBER Working Paper No. 17632. Issued Dec. 2011.

[11] Magesh. S, Nimala. K "A survey on machine learning approaches to social media analytics" International Journal of Applied Engineering Research, Volume 11, Number 4 (2016) pp 2411-2416.

[12] www.ibmbigdatahub. com/infographic/four-vs-bigdata dated 11-03-2018.

[13] Tania Cerquitelli, Silvia Chiusano, MirkoKampf, "Special session on big data: new trends and applications", Big Data, 2013.

[14] www.Expert system. com/machine-learning-definition/, dated 11-03-2018.

[15] Lloyd Allison, "Types and classes of machine learning and data mining", Proceedings of the 26th Australasian Computer Science Conference, Vol. 16, Page No. 207-215, 2003.

[16] Junfei Qiu, Qihui Wu, Guoru Ding, Yuhua Xu and Shuo Feng, "A survey of machine learning for big data processing", EURASIP Journal on Advances in Signal Processing (2016)2016:67.

[17] K Slavakis, GB Giannakis, G Mateos, "Modeling and optimization for big data analytics: (statistical) learning tools for our era of data deluge" IEEE Signal Proc Mag 31(5), 18–31 (2014).

[18] www. blog. leanix. net/en/what-is-the-difference-between-artificial-intelligence-and-machine-learning, dated 14-03-2018.

[19] Hammoudeh, Ahmad, "A Concise Introduction to Reinforcement Learning", 10.13140/RG.2.2.31027.53285 (2018).

[20] JiaLiu, MaoguoGong, QiguangMiao, "Modeling Hebb Learning Rule for Unsupervised Learning", Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17).

[21] Erik G. Learned-Miller, "Introduction to Supervised Learning", February 17, 2014.

[22] Kim, P. "The Forrester Wave: Brand Monitoring, Q3 2006" Forrester Wave (white paper), 2006.

[23] Pang, B., "Using very simple statistics for review search: An exploration", In: Proceedings of the International Conference on Computational Linguistics (COLING) (Poster paper), 2008.

[24] Ramakrishnan, N., Butler, P., Muthiah, S., "Beating the news with embers: forecasting civil unrest using open source indicators", In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2014, pp. 1799–1808. ACM, New York (2014).

[25] Swati Agarwal, Ashish Sureka, and Vikram Goyal "Open Source Social Media Analytics for Intelligence and Security Informatics Applications", Big Data Analytics (Springer), 4th International Conference, BDA 2015, Hyderabad, India, December 15-18, 2015.

[26] Agarwal, S., Sureka, A., "A focused crawler for mining hate and extremism promoting videos on YouTube", In: 25th ACM Conference on Hypertext and Social Media (HT), pp. 294–296 (2014).

[27] Rongjing Xiang, Jennifer Neville and Monica Rogati "Modeling relationship strength in online social networks". In Proceedings of the 19th international conference on World wide web, pages 981-990. ACM, 2010.

[28] P. Singla and M. Richardson. Yes, there is a correlation: - from social networks to personal behavior on the web. In WWW '08, 2008.

[29] Hatzivassiloglou, V., McKeown, K., "Predicting the Semantic Orientation of Adjectives. In: Proc. 8th Conf. on European chapter of the Association for Computational Linguistics", Morristown, NJ, USA, Association for Computational Linguistics 174 – 181, 1997.

[30] Dave, KL., Pennock, D., "Mining the peanut gallery: Opinion Extraction and Semantic Classification of Product Reviews", In: Proceedings of WWW 519-528, 2003.

[31] Pang, B. and L. Lee, Vaithyanathan, S., "Thumbs up? Sentiment classification using machine learning techniques", In: Proceedings of Conference on Empirical methods in natural Language Processing (EMNLP), Philadelphia, July 2002, 79 - 86. Association for Computational Linguistics, 2002.

[32] Turney, P., "Thumbs Up or Thumbs Down? Semantic orientation applied to unsupervised Classification of Reviews," In: Proceedings of the Association for Computational Linguistics (ACL), pp. 417–424, 2002.

[33] Jonathan Magnusson, "Social Network Analysis Using Big Data technology", Uppsala Universitet, January 2012.

[34] Agreste, S, De Meo, P, Ferrara, E, Piccolo, S and Provetti, A., "Analysis of heterogeneous social network of humans and cultural objects", IEEE Transactions on Systems, Man and Cybernetics: Systems, 45 (4), 559 – 570 (2015). http://dx.doi.org /10.1109/TSMC.2014.2378215.

[35] Devakunchari R, Valliyammai C, "Big social data analytics: opportunities, challenges and implications on society", International Conference on Communication, Media, Technology and Design 27 - 29 May 2016 Zagreb – Croatia.

[36] Al-Jarrah, O.Y, Yoo, P.D, Muhaidat, S, Karagiannidis, G.K, Taha, K, "Efficient Machine Learning for Big Data: A Review", Science Direct-Elsevier, Big Data, Analytics, and High-Performance Computing, 2 (3), 87–93 (2015).

[37] Borko Furht, "Handbook of Social Network Technologies and Applications", New York, Springer. http://dx.doi.org /10.1007/978-1-4419-7142-5 (2010).

[38] Arun Ganesh., "Crowdsourcing flood data for Chennai [Blog Post]", Retrieved from: https://www.mapbox.com / blog /chennai-flood-map/ (2015, Dec.2).

[39] Venner, Jason. Pro Hadoop. Apress. ISBN 978-1-4302-1942-2 (2009).

[40] Web content available on the link: ―http://www.sas.com/en_us/insights/big-data/what-is-bigdata.html‖ on the dated: 16-08-2015.

[41] Web content available on the link: ―http://www-01.ibm.com/software/data/bigdata/what-is-bigdata.html on the dated: 16-08-2015.

[42] Web content available on the link: ―http://searchbusinessanalytics.techtarget.com/definition/big-dataanalytics‖ on the dated: 16-08-2015.

[43] Jai Prakash Verma, Smita Agrawal, Bankim Patel, Atul Patel, "Big data analytics: challenges and applications for text, audio, video, and social media data", International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.5, No.1, February 2016.

[44] Amir Gandomi , Murtaza Haider, " Beyond the hype: Big data concepts, methods, and analytics", International Journal of Information Management 35 (2015) 137–144.

[45] C.L. Philip Chen, Chun-Yang Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", Contents lists available at ScienceDirect Information Sciences, 275 (2014) 314–347.

[46] Stephen Kaisler, Frank Armour, J. Alberto Espinosa, William Money, "Big data: issues and challenges moving forward", 2013 46th Hawaii International Conference on System Scien 15301605/12, IEEE.

[47] Edmon Begoli, James Horey, "Design principles for effective knowledge discovery from big data", 2012 Joint Working Conference on Software Architecture & 6th European Conference on Software Architecture, 978-0-7695-4827-2/12, IEEE.

[48] Yang Song, Gabriel Alatorre, Nagapramod Mandagere, and Aameek Singh, 2013, IEEE International Congress on Big Data‖, 978-0-7695-5006-0/13, IEEE.

[49] Charu. C. Aggarwal, "Social network data analytics", IBM Watson Research Center, US.

[50] Ostrowski D. A, "Map reduce design patterns for social networking analysis" Page nos. 316-319, IEEE International conference on Semantic Computing, 2014.

[51] Anteneh Ayanso, "Harnessing the power of Social Media and web Analytics", IGI Global, February 2014.

[52] https://en. wikipedia. org/wiki/Graph –Database.
[53] William M. Campbell, Charlie K. Dugli, Clifford J. Weinstein, "Social network analysis with content and graphs", Lincoln Laboratory Journal, Page Nos 62-81, Vol 20 No. 1, 2013.
[54] K. L. Du, M. N. S. Swamy, "Neural networks and statistical learning" Springer-Verlag London, 2014.
[55] Prem Melville, Vikas Sindhwani, Richard D. Lawrence, Estepan Meliksetia, "Machine Learning for Social Media Analytics".
[56] Steve Oberlin, "Machine Learning Cognition and Big Data", CA Technologies, 2012.
[57] www.towardsdatascience.com/how-to-make-sense-of-social-media-using-machine-learning-8a3db1506d03 dated 25-04-2018.
[58] Ms. Priyanka Patel, Ms. Khushali Mistry, "A Review: Text Classification on Social Media Data", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 17, Issue 1, Ver. IV (Jan – Feb. 2015), PP 80-84.
[59] Amir Atiya, Mohammed Aly, Mahmoud Nabil, "Social Media Analytics", PRP Project, Faculty of Engineering, Cairo University, 2015.
[60] Yee Ling Phua, "Social Media Sentiment Analysis and Topic detection for Singapore English", Thesis, Calhourn Postgraduate school, September 2013.
[61] https://en. wikipedia. org/wiki/ID3-algorithm.
[62] B. Hassina, A. Merbarha, H. EZZ, Kouri, M. ERRITALI, "A Comparative study of decision tree ID3 and C4. 5", International Journal of Advanced Computer Science and Applications special Issue on Advances in Vehicular Ad Hoc networking and Applications pp-13-19.
[63] K. Nirmala, S. Satheesh kumar and Dr. J. Vellingiri, "A Survey on Text categorization in Online Social Networks", International Journal of Emerging Technology and Advanced Engineering Volume 3, Issue 9, September 2013.
[64] www. en.wikipedia. org /wiki/Support_vector_machine dated 25-4-2018.
[65] Jing Bai, Jian-Yun Nie, "Using Language Models for Text Classification".
[66] J. Holton Wilson, "An analytical Approach to detecting Insurance Fraud Using Logistics Regression", Journal of Finance and Accountancy.
[67] M. M. Abo Khedra, A. A. Abd El-Aziz, Hesham A. Hefny, "Social Network Analysis through Big Data Platform Review", 2019 International Conference on Computer and Information Sciences (2019 ICCIS) 978-1-5386-8125-1©2019 IEEE.
[68] Saeed, Z., Abbasi, R.A., Maqbool, O. *et al.,* "What's Happening Around the World? A Survey and Framework on Event Detection Techniques on Twitter". *J Grid Computing* **17,** 279–312 (2019). https://doi.org /10.1007/s10723-019-09482-2.
[69] M. K. Hayat *et al*., "Towards Deep Learning Prospects: Insights for Social Media Analytics," in *IEEE Access*, vol. 7, pp. 36958-36979, 2019.
[70] Khan, Hikmat Ullah, et al. "Modelling to identify influential bloggers in the blogosphere: A survey", *Computers in Human Behavior* 68 (2017): 64-82.
[71] Masood, Ali & Abbasi, Rabeeh & Maqbool, Onaiza & Mushtaq, Mubashar & Aljohani, Naif & Daud, Ali & Aslam, Muhammad & Alowibdi, Jalal. "MFS-LDA: a multi-feature space tag recommendation model for cold start problem.", (2017) Program. 51. 00-00. 10.1108/PROG-01-2017-0002.
[72] Muhammad Aslam Jarwar, Rabeeh Ayaz Abbasi, Mubashar Mushtaq, Onaiza Maqbool, Naif R. Aljohani, Ali Daud, Jalal S. Alowibdi, J.R. Cano, S. García, Ilyoung Chong, "CommuniMents: A Framework for Detecting Community based Sentiments for Events", International Journal on Semantic Web and Information Systems (IJSWIS), 13(2): 87-108, 2017.
[73] Haneef F., Abbasi R.A., Sindhu M.A., Khattak A.S., Noor M.N., Aljohani N.R., Daud A. & Arafat S., "Using network science to understand the link between subjects and professions", Computers in Human Behavior (2020), doi: https : // doi.org/10.1016/j.chb.2019.106228.
[74] Rabeeh Ayaz Abbasi, Onaiza Maqbool, Mubashar Mushtaq, Naif R. Aljohani, Ali Daud, Jalal S. Alowibdi, Basit Shahzad. Saving Lives, "Using Social Media: Analysis of the Role of Twitter for Personal Blood Donation Requests and Dissemination. Telematics and Informatics", 35(4): 892-912, 2018.
[75] Maria Óskarsdóttir, Cristián Bravo, Wouter Verbekee, Carlos Sarrautef, Bart Baesens, Jan Vanthienen, "Social Network Analytics for Churn Prediction in Telcom Model Building Evaluation and Network Architecture", Expert Systems with Applications (2020) arXiv:2001.06701 [cs.SI]
[76] https : // www. Netbase . com / blog/ free-social-media-analytics-tools/ dated 31st March, 2019.
[77] S. Stieglitz, M. Mirbabaie, Björn Ross, C. Neuberger,"Social media analytics – Challenges in topic discovery, data collection, and data preparation", International Journal of Information Management, Volume 39, April 2018, Pages 156-168.