

An Overview of Social Media Analytics

*T. Lakshmi Praveena*¹
*Research Scholar*¹,
SPMVV,Tirupathi

*Dr. N. V. Muthu Lakshmi*²
*Assistant Professor*², *Dept of CS*²
SPMVV,Tirupathi

Abstract-Social media is the medium used to share views, ideas and opinions related to the current environment or things happened. Social media is used in different areas like education, politics, shopping, marketing, sports, health care etc. The information gathered from social media sites allows deriving learning information. The data derived from Social media is very huge also termed as big data, measured in billions and is being successfully used for analysis to predict uncover information. The results obtained from analysis over social media finds interesting values and relationships between data which helps to make right strategic decisions. Now a days, finding efficient methods to do analysis over social media data is becoming an important emerging research topic for various people especially for business people. In this paper, evolution of social media, types of users, benefits of social media are discussed. Today, many social media are being used effectively and also these social media data help the people to know hidden information by analysis. Data analysis plays an important role in social computational science which helps the people in decision making. The data analysis over social media of big data is also called social media data analytics or simply social media analytics and is giving huge opportunities for finding new ideas to the analyst to understand the emerging trends. This paper also addresses different types of analysis over social media, the popularly used tools for social media analytics. The various research challenges in analyzing social media are also presented.

Keywords:

Data analysis, Data Analytics, Social Media Analytics, Big Data

I. INTRODUCTION

a. What is social media?

Social media is a communication media between the users. Social media is the medium used to share views, ideas and opinions related to the current environment or things happened. Walter and Riviera defined social media as “Relationship that exist between network of people” in 2004. *Social media* is also known as “anywhere – anytime” available information. Social media does not require any point to point communication so irrespective of the receiver information can be send as broadcast message. Merriam and Webster defined social media as “*Social media* is the form of electronic communication through which users create online communities to share information, ideas, personal messages, and other content”. As per the Lisa Buyer “*Social media* is transparent and interactive form of public relations. It combines the real time content with authentic peer-to-peer communication”. Michelle Chmielewski definition of social media is “*Social media* is about what we do and say together, worldwide, to communicate in all directions at any time, by any possible digital means.

II. EVALUTION OF SOCIAL MEDIA

The earliest methods used for communicating across great distances were written correspondence delivered by hand from one person to another. The telegraph was invented in 1792 and is allowed messages to be delivered over a long distance faster than traditional method such as horse and rider could carry them. In 1890 the telephone is invented

and in 1891 the radio is invented to communicate across long distances instantaneously.

In 20th Century super computer is invented and this lead to create networks for communication which lead to birth of internet. The initial form of the Internet, such as CompuServe, was developed in 1960s and earlier form of email was developed in this time. Networking technology had improved by 70’s and UseNet allowed users to communicate through a virtual newsletter. From 1980 home computers become more common and social media became more sophisticated. The first social media web site is Six Degrees was created in 1997. It enabled users to upload a profile and make friends with other users. The first blog site became popular in 1999 and created a social media sensation that’s still popular today.

The invention of blogging, social media began populared, example sites are MySpace and LinkedIn, Photobucket and Flickr etc. In the early years of 2000 YouTube, Facebook and Twitter are started and became available to users. Other sites like Tumblr, Spotify, Foursquare and Pinterest began popping up for media information.

2.1 Users of Social Media

The users of social media are categorized based on 2 properties, one is based on usage of social media and second is based on addiction to social media. These are categorized into different types. There are different users which are using social media regularly. This user information can be used for decision making.

- a. *Social Butterflies*: The Mix-n-Minglers are users of social media who Identifies, Understands and Influences social Media. These users can be of different types based on the activity performed by the user.
- b. *Never there*: These users use the social media to post their profiles and collects the different users digital data.
- c. *Informers*: These users are announcers or informers of social media. These users spreads the reviews and views of website which can improve business or sometimes it can spoil the business.
- d. *Unsatisfied Customers*: These users are different users which have negative opinions on site and post the negative feedback which may spoil the business. So these users satisfaction defined the company brand.

There are four types of users based on addiction to social media

- 1. *The Ultras*: These users check feeds for number of times a day and Ultras use social media only for this purpose.
- 2. *The Dippers*: These users are rare users who access their pages in social media websites.
- 3. *The Deniers*: *Daniers* maintain social media but not active. These users feel anxious when there is interruption in accessing network.
- 4. *The virgins*: These are active users and take tentative steps in social media websites.

III. BENEFITS OF SOCIAL MEDIA

Social media has number of benefits. It is benefitted to different classes of people in different ways. Young people use social media to gather more information from others by making social connections with the people. Business people use social media for improving the business and to analysis the customer requirements.

- a. Social media As Communication Media: Social media of became communication alternative to phone calls. It provides broadcast messaging between users of similar features or properties.
- b. Social media for information Sharing: Social media platform is commonly used for sharing information like photos,events.
- c. Social Media As Research Media: Social media is benefited for professionals and academics for sharing research ideas. This improves professional reputation and those looking to increase a professional reputation.
- d. Social Media in Business Marketing: Social media solves business marketing problems by using customer responses and analysis on customer data. Social media allow businesses to reach a large number of potential customers. Social media provides a way to address the

questions, comments and concerns of existing customers, building brand loyalty. Social media networks or communities are discussed in next section.

IV. SOCIAL MEDIA COMMUNITIES

Social media communities are types of networks used to establish relationship between users or customers of users. These are used to share opinions and views. For ex PM Modi announced demonetization in November which spread instantly to all people or users through social media networks. For ex in case of natural disasters social media plays an important role in helping public or in gathering public. These social media networks arle categorized into different types.

- a. *Relationship Networks*
Relationship networks provides unique chance for brands to connect to their users on a personal level. These days, it is necessary for most brands to have a Facebook Page or a Twitter account, in order to reach out to their audience online and answer any customer service queries that may arise.
- b. *Media Sharing Networks*
This type of social network is defined by the primary type of media shared among users.
- c. *Online Reviews*
Location-based review services such as Yelp and Urbanspoon are social networks adopt geolocation.
- d. *Discussion Forum*
Discussion forums are the oldest social network which is used to share ideas and solve problems.
- e. *Social Publishing Platforms*
Social publishing platforms consist of blogs and microblogs, where long and short messages are shared with other users.
- f. *Bookmarking Sites*
Bookmarking sites are used to search for the content when we have a vague idea of content. These are web services like StumbleUpon, Pinterest, and Flipboard, where users collect content from the Internet, and save it to their account on the platform. This content can be private or public, and shared with other users. These bookmarking sites will then suggest content similar to the links or images related to the content saved on the network.
- g. *Interest Based Networks*
These networks are used to find the people with the common interests or opinions.
- h. *E-commerce Networks*
E-commerce is the present trend of social media where users can watch and purchase goods with a single click.
- i. *Social News Networks*

Services that allow people to post various news items or links to outside articles and then allows it's users to "vote" on the items.

These are the different social media communities that are used for different purposes for sharing different types of contents. The next section gives overview of Data Analytics.

V. DATA ANALYTICS

Data analysis is the process of applying logical and statistical techniques to evaluate and describe raw data to extract useful information. Different forms of data analysis include text information, graphs and charts. Data analysis methods are used to refine raw data and derive information.

5.1 Methods of Data Analysis

The different methods used for data analysis are One Way ANOVA, Multiple Regression, Paired Sample T-test, and Logistic Regression.

- j. Multiple regression is the standard method used for data analysis. This method is used to examine independent variables and dependent variables. Multiple regression analysis method is used when the value of one variable is predicted from one or more other variables.
- k. Logistic regression is used to examine non linear variables. This regression is used when the variable is a binary variable. This analysis is predictive analysis. Logistic regression is used to find relationship between dependent or binary variable and independent or interval variable.
- l. One way ANOVA is used to test variances and to test two or more samples. The one-way ANOVA compares the means between the groups. And determines whether these means are statistically significantly different from each other. Specifically, it tests the null hypothesis:

$$H_0=U_1+U_2+U_3+.....U_k$$

where U = group mean and k = number of groups. The one-way ANOVA returns a statistically significant result of the alternative hypothesis which is two groups mean that are statistically significantly different from each other.

- m. A paired sample t-test is used to compare two population means. This also known as dependent sample t test which is a statistical procedure used to find the mean difference between the two sets is zero.

VI. DATA ANALYTICS IN SOCIAL MEDIA

Social media data is analyzed to store data sets and create presentations and these visualizations are accessible by other users. Social media data is analysed to predict interesting values and find relationships between data. The

statistical methods, machine learning methods and data mining methods are used for data analysis.Types of social media data analysis are as follows

- a. *Sophisticated data analysis* : based on context, content and sentiment.
- b. *Time Consideration* : based on days or hours
- c. *Influence analysis* : based on impact of news or messages to check quality of data.
- d. *Network Analysis* : based on data grows or migrates through the network.

Different domains that requires social media data analysis for decision making and to improve the performance of company. Maximum 90% of business domains are using social media data analysis for their business, some of these are:

- a. *Bank domain* uses social media analysis to market products and services. Social media analysis shows impact on risk management, product design, business forecasting and competitive analysis.
- b. *Education domain* uses social media analysis to exchange new analytical methods and to discuss new problems raised in research. Students uses to learn new technologies.
- c. *Consumer products domain* uses social media analysis to analyze social media to encourage new innovations and improve product design or service.
- d. *Entertainment and communication domain* uses social media analysis to gather the views, comments, opinions or feedbacks about the announcements or show.

Necessity of data analytics in social media is to analyze the followers, to identify influencers and to compare different platforms and products. Data analysis in social media is also used to simplify data from dozens of networks, millions of people and a variety of activities. Used to understand the attitudes and preferences of customers or users.

VII. TOOLS FOR SOCIAL MEDIA ANALYTICS

Architecture of Social Media Analytics and lifecycle of social media analytics is represented in different phases. These are different steps in analyzing the data and presenting the data. The data may be online or offline data. The data collected from different social media sources, services and tools. This data is preprocessed for analysis. The resulted data is interactive data and is visualized using visualization tools.

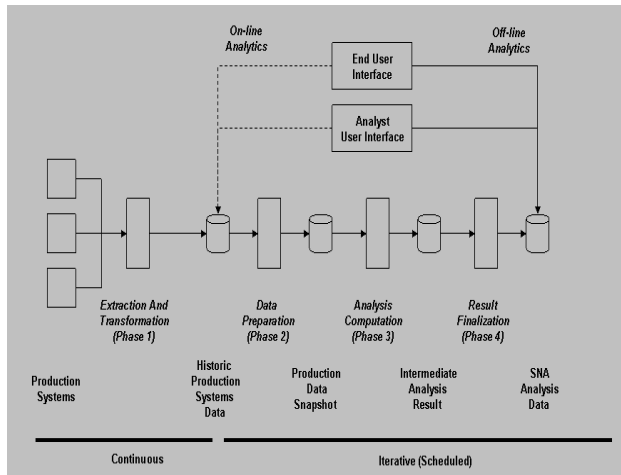


Fig 1. Lifecycle model of SMA

Different phases of SMA lifecycle and tools required are:

1. *Extraction and Transformation*

Data collected from the production systems. These systems are the web interfaces or web sites. This data is continuous data which may be in the form of numerical or character. The tools used in this phase are

- Transformation Tools:
 - Google Fusion Tables
 - Zoho Reports
 - IBM’s Many eyes
- Extraction Tools
 - Google trends
 - Social Mention
 - Twitter Feeds
 - Social Pointer

2. *Data Preparation*

Collected data is prepared for the analysis. This phase uses the preprocessing tools to clean the data and stores the data. Relational databases and data warehouses are used to store the data. The data collected is historic data from multiple social media sources.

- Data cleaning Tools
 - Google Refine
 - Data Wrangler

3. *Analysis Computation*

The data that collected and cleaned is used for analysis. The analysis is performed based on analytics dash boards and programming languages like MATLAB, Java, Python. The analysis tools used are

- Gephy
- Excel plugin Node XL

Social Media Analytics Techniques Used are Computational Science methods. Machine learning is a computation science method which is divided into Supervised and unsupervised.

Supervised techniques are Classification and Regression. The Classification techniques used are SPM, Naïve Bayes, Nearest Neighbourhood method. The Regression methods are Neural Networks and Decision Trees. Regression is of 3 types linear, non linear and non parametric.

Unsupervised techniques used are clustering. Different types of clustering methods are K-Means, Heirarchical, Neural NWs, Gaussian and Hidden Markov analysis.

4. *Result Finalization*

The analysed data is presented in the form of graphs, charts. Data visualization tools are used for this phase. It provides interactive result of SMA. The tools used for visualization are

- Apache Hadoop
- Amazon Kinesis
- SAS Visual Analytics
- Tableau

VIII. REAEARCH CHALLENGES

- *Scrapping* data from different sources of social media.
- *Data Cleansing* is preprocessing the collected data by removing noise data and un-necessary data.
- *Holistic Data Sources* is the availability of data from large social media sources. It requires permission to access data.
- *Data protection* is the main factor that vendors of social media web sites or companies reject to share data. The protection to be provided.
- *Data analysis* plays an important role in social computational science which helps in decision making.
- *Analytics Dashboards* are used to show the analysis results of social media analytics.
- *Data Visualization* is used to present the analysis results with different tools.

CONCLUSION

The social media gave rise to a new set of analytics to find the consumer preferences. So is used for marketing campaigns. The social media analytics is giving huge opportunities for innovative analytics needed to understand emerging trends in social media. Social media analysis creates different research challenges in analytics. Beacause of the requirement to perform sentiment analysis, opinion mining, machine learning etc. Paper discussed on the introduction to social media and social media analytics. The different methods in data analytics and tools used in data analytics. And also given research challenges available in social media analytics.

REFERENCES

- [1]. Gitanjali Kalia,A Research Paper on Social media: An Innovative Educational Tool, Chitkara University, Punjab
- [2]. Cioffi-Revilla C (2010), Computational social science. Wiley Interdiscip Rev Comput Statistics
- [3]. Galas M, Brown D, Treleaven P (2012), A computational social science environment for financial/economic experiments. In: Proceedings of the Computational Social Science Society of the Americas
- [4]. Khan A, Baharudin B, Lee LH, Khan K (2010), A review of machine learning algorithms for text-documents classification. J Adv Inf Technol
- [5]. Lazer D et al (2009) Computational social science.
- [6]. Ana Mihanovic, Hrvoje Gabelica, ZivkoKrstic 2014 Big Data and Sentiment Analysis using Knime: Online Reviews Vs. Social Media, MIPRO Opatija, Croatia.
- [7]. Bogdon Batrinca, Philip C. Treleaven 2014 Social media analytics: a survey of techniques, tools and platform, Department of Computer Science, Gower Street, London, UK published in Springer.
- [8]. Umar Ruhi,2012, Social media analytics as a Business intelligence Practice : Current landscape & Future Prospects,Journal of internet social networking and virtual communities
- [9]. Yafeng Lu, Robert Krüger, Student Member, IEEE, Dennis Thom, Feng Wang, Steffen Koch, Member, IEEE, Thomas Ertl, Member, IEEE, and Ross Maciejewski, Member, IEEE,2014, Integrating predictive analytics and social media
- [10]. ZN Gastelum, KM Whattam, 2013, State of the art of social media analytics research.