

Social Media Analytics using Visualization

Shweta Patnaik¹, Swati Sucharita Barik²

¹(Computer Science & Engineering, Centurion University, Bhubaneswar
Email: 160303110004@cutm.ac.in, shwetapatnaikplk@gmail.com)

²(Computer Science & Engineering, Centurion University, Bhubaneswar
Email: swati@cutm.ac.in)

ABSTRACT

In this current world bigdata analytics is playing a very crucial role in business, Government communities, social media, health care and in many different platforms. More than 1.5 billion users worldwide use social media, in the form of real time, interactive communications made available through blogs, tweets, updates, images and videos. Because people in the society share their thoughts and preferences in the social media, there is an opportunity to analyze the social media data.

Many challenges will come ahead in the development of social media analytics. This technology is used to improve the accuracy, performance and scalability as social media is growing very rapidly. This paper discusses the work of visualization and analysis methods and their result.

Keywords - Data visualization, social media data, social media analytics, Tableau.

I. INTRODUCTION

The current era has a rapid growth in information technology. As a result, it produces a huge number of data. Approximately 30 percent and more in storage capacity, network bandwidth, and CPU processing power have been used to collect, store, transmit, and process unstructured and noisy data.

Now a day's millions of people are using social media. Which itself is dynamic, distributed and a huge network. The user generated contents are of different formats like- video, text, images and some geographical data. As we can see that, the social media data is high in volume, and there may be some redundant contents. It creates a challenge to analyze those huge amounts of data having high frequency.

According to Ahrens et al (2001), the size of dataset in data visualization tasks will continue to grow because

of its ever-increasing applicability in various domains [1]. It is obvious that the influence and significance of data visualization cannot be underestimated [1].

Social Media Analytics is all about collecting data produced from Social Media platforms like Facebook, Twitter, LinkedIn, WhatsApp, Wikipedia, YouTube, Pinterest, Instagram, Tumblr, Snapchat, Google+, WeChat, and many others. The data are in the form of comments, tweets, posts, likes, shares and links, geographical data, microblogs for fetching the information from the dataset using social media analytics, to make the right business decisions. Social Media Analytics also includes development and calculation tools and frameworks which collect, monitor, analyze, summarize, and visualize social media data. It is also used for learning customer sentiment and behavior and is also useful to get the insights on product reviews and data for creating better marketing strategies and improved customer service.

The general approach for social media analytics starts by identifying the credible and mineable source of information from the diverse social media platforms used for Microblogging, Blogging, Community-based Question Answering (CQA), Chats, Forums, Media Sharing, and Hybrid Applications which generate huge volumes of noisy, distributed, unstructured and dynamic data [2].

1.1. SENTIMENT ANALYSIS

Sentiment analysis states to the use of natural language processing (NLP), text analytics, computational linguistics, and biometrics in identifying, extracting, quantifying and studying the affective states and subjective information [3] to capture the emotion of the speaker or writer.

II. DATA VISUALIZATION

Data visualization is a general term that describes any effort to help people understand the significance of data by placing it in a visual context. Patterns, trends and correlations that might go undetected in text-based data can be exposed and recognized.

It is important because, by using charts or graphs visualizing large amount of data is easier than transferring them to spreadsheets. Data visualization is a quick, and an easy way to describe concepts in a universal manner, and we can do experiment with different situations by making some adjustments. It can also identify areas that needs attention or improvement, simplify the factors that influence customer behavior, helps us to understand which products to place where, predict sales volumes.

2.1. Comprehend information quickly:

By using graphical or pictorial representations of business information, businesses are able to look for large amounts of data in clear, and in organized ways, and give conclusions from that information.

2.2. Identify relationships and patterns:

The wide amounts of complicated data make sense when presented graphically. Businesses can identify parameters that are highly correlated. Some of the relationships will be clear, but others won't. Identifying those relationships helps the organizations to focus on those areas which are most likely to influence their most goals.

Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments [2].

2.3. Pinpoint emerging trends:

Using data visualization to discover trends – both in the business and in the market – can give businesses an edge over the competition, and ultimately affect the bottom line. It's easy to spot outliers that affect product quality or customer churn, and address issues before they become bigger problems [2].

2.4. Communicate the story to others:

Once a business has uncovered new insights from visual analytics, the next step is to communicate those insights to others. Using charts, graphs or other visually impactful representations of data is important in this step because it's engaging and gets the message across quickly [2].

Data visualization can also:

- Identify areas that need attention or improvement.
- Clarify which factors influence customer behavior.
- Help you understand which products to place where.
- Predict sales volumes [2].

2.5. Different Visualization Tools:

2.5.1. Tableau:

Tableau is known as the prime data visualization software. It is one of the best data visualization platforms, it's simple, drag and drop based interface gets the major point. Founded in 2003, Tableau continues to be the market leader given a wide variety of visualizations and efficient finish that it offers. It is offered on desktop, server and cloud versions and can connect to local and remote data and also to live data. We will discuss more about Tableau later.

2.5.2. SAP Lumira:

SAP Business Object Lumira is a self-service visualization tool used to connect, visualize, discover and communicated based on the large data sets. It brings data exported from sources such as Excel and others to life by creating beautiful, interactive maps and infographics. Its dashboard performs the visual BI analysis and shares the insights with the team.

2.5.3. SAS Visual Analytics:

It is one of the most traditional vendors in advanced analytics space, SAS visual analytics has been in the industry for a long time offering insights to various businesses. It is capable of delivering fast answers to complex questions regardless of the size of the data. It processes information quickly, identifying relationships and pattern and pointing trends based on the data. It's used across industries like healthcare, digital advertising etc. Its web-based, interactive data exploration mode suits the needs of all types of users.

2.5.4. Qlik:

It is one of the Leaders in Gartner's 2017 Report, it offers a wide range of products that provides data visualization. QlikSense, QlikView, Qlik Analytics Platform are some of its products. It lets you search and explore vast amounts of data and to analyze, reveal, collaborate and act based on the insights. QlikSense allows you to explore the data intuitively and creating smart visualizations with drag-and-drop ease. It helps

you explore complex data like never before. Its available in various versions for desktop, server, and cloud.

2.5.5. Oracle Visual Analyzer:

Claiming a place at the Magic Quadrant Business Intelligence and Analytics Platform report by Gartner this year after last year's loss, this 2015 introduced web-based tool is provided within the Oracle Business Intelligence Cloud Service. Apart from providing advanced analysis and interactive visuals, Oracle Visual Analyzer also gives a configurable dashboard. It's mostly preferred by organizations with large scale enterprise deployments where it can help create detailed analyses and carefully curated dashboards resulting in detailed reports.

III. SOCIAL MEDIA DATA

Social media data defines all of the raw information collected from individual social media activity. Social media data tracks how individuals participate with your content or channels like LinkedIn, Facebook, and Twitter. It gathers numbers, percentages, and statistics from which you can gather the performance of your social media strategy. It's best to view social media as a source of raw data, which includes the metrics like shares, likes, conversions, comments, mentions etc.

Social media data are your raw information. With the social media analytics, we can answer number of questions about the success of your social media activities, such as:

- Which networks contribute the most to lead generation?
- What types of content makes audiences click, share, and convert?
- What are my top-converting posts?

3.1. How to collect social Media Data:

Social media data is more than likes and shares. It provides B2B marketers a deeper insight into who your audiences are and how they engage with your company. Collecting social media data is important for making confident business decisions and improving the processes of other departments. Social media analytics gives you raw data and statistical insights to understand the success of every post, campaign, and network. Beyond the measures of likes, shares, and impressions, it enables you to track clicks and leads from

social media activities, enabling you to identify which channels and messaging drive revenue for your business.

By collecting social media data, we can integrate this data with the entire marketing ecosystem – including your marketing automation, CRM, and BI platforms. These powerful integrations allow you to connect the dots of your buyer journey so you can deliver a more engaging experience across every channel and department.

For instance, Facebook data:

We can collect the facebook data by using Facebook's Graph API apart from other built-in tools to scrape Facebook data [2].

- The Graph API is the primary way to extract data in Facebook platform. Facebook Graph API allows your application program to read and write Facebook Social Graph. It provides data in the form of text as well as graphs. Built-in libraries for Facebook [4].
- We can make use of Graph API and scrape data using python [4].
- Rfacebook: The package Rfacebook lets you to access Facebook data using R programming [4].
- PHP library: The facebook SDK is used for text and graphs from facebook using PHP for web page development [4].
- Apache Flume: We can configure Apache Flume to automatically gather Facebook data using Graph API [2].

3.2. Types of Social Media Data:

Data plays a significant role for identification, resource allocation and content programming. So social media data has been categorized, such as:

3.2.1. Performance data:

Performance data are simple and detailed, as required. Generally, it gives a quick, and easy overview to manage social channels. It can also build its own custom reports which gives multi-channel project critical data and insights without any interruptions. These data includes click through rates, document downloads and other KPI data depending on the project.

3.2.2. Conversation data:

It mines the real-time conversations, brand mentions, user characteristics, sentiment and content. By analyzing this, the analyst will be able to make cross-channel connections, cross-reference data and apply filters to separate trends. There are a few enterprise level

analytics and monitoring applications being used for this and fetch the data directly from the social media account or any other network API's. By observing to real-time conversation data allows a company to respond quickly, fulfilling and sharing positive stories or taking steps to remedy service failures, solve problems and learn about their customers feelings.

3.2.3. Industry or topic data:

By monitoring industry and topic data, analyst can identify key performance indicator (KPI). It will be helpful to measure a brand's share of voice within the industry and monitor opponent's activity. It also involves tracking hashtags, keywords and phrases people use to relate to a question, issue or event of a brand. Tracking industry and topic data helps us listen to a wider knowledge.

3.2.4. Traffic and link data:

These kinds of data are mostly generated by social media pages, ads, and blogs. Some of the companies use their website as the central hub for digital content, communications and resources. Social channels, SEO, ad campaigns and the like all generate traffic data which leads back to the website. It could be a campaign landing page, video, PDF, blog post, support knowledge base or a contact form as a few examples. The analyst has to understand, where traffic comes from, which social networks, items of content or ads are adding the most value.

With link tracking we can start developing stories about where your content is shared, published and accessed. We can watch as a resource is shared across multiple social channels, who the key influencers are who shared it and who generated the most re-shares of the content online. Digging deeper we can learn demographic, location and sentiment trends related to each item of content or link we share.

3.2.5. Conversion or results data:

Some companies easily regulate the value of a visit to their website based on average conversion rates and the dollar value of purchases. In other case the ticket purchases from a Facebook ads or campaign tab through to the ticket booking page and can track the interaction through to the confirmation page. It can be tracked that how many times a post has been viewed or downloaded and who are the visitors. Then there might be a cross reference that information with following conversation data like hash-tagged conversations, quote tweets from

the doc or other discussions that resulted from the interaction.

3.2.6. Referrals Data:

Ratings, reviews and nonverbal attitudes generate referral data. Some applications are:

- A radio station in Australia, "triple j" uses its Twitter data to understand whether listeners like their programs or not by monitoring the rate of re-tweets and responses (Schlagwein, 2014)
- IKEA uses interactive catalogue to engage potential customers who share posts of relevant products within their social media circle. It resulted into a big spike into its sales figure (Scott, 2013).
- McDonald's Japan promotes new products through utilizing fan base in blog and Twitter. These fans talk about their preferences of foods and the company and occasionally receive offers of free food (Edelman & Salsberg, 2010) [4].

3.2.7. Local Data:

Real time location data of consumer. Some application:

- Using social media, Levi Strauss offered location specific deals by generating word-of-mouth by their customers. And in one particular case, 400 customers' word-of-mouth brought 1600 people to the stores (Muñoz & Strotmeyer, 2010) [4].

3.2.8. Intentional Data:

Data in social media that indicate consumers' future buying intentions and activity. Some applications are:

- By semantic analysis of Facebook status updates, Shift (an American company) measures people's purchase intent and sharing the data with their client companies as service (Beckland, 2011).
- Hornall Anderson, a marketing company, generates purchase intention score for their client companies through analyzing social media data; clients include a big number of retail and consumer companies (Gleanster, 2013).
- By analyzing intention data, companies can design better offer and plan for consumers, better example can be set for real estate, bank, insurance etc [4].

IV. SOCIAL MEDIA DATA ANALYTICS

Social Media Analytics deals with development and evaluation of tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data.

- Facilitate conversation and interaction between online users.
- Extracts useful patterns and information [5]

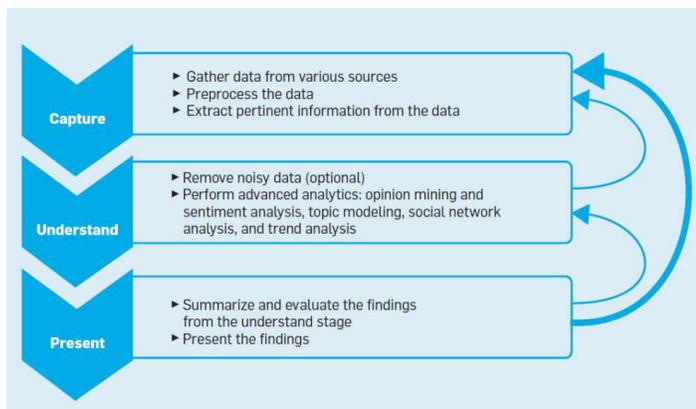


Fig:1 Social Media Analytics Process

Social media analytics provides a human trace to the social scientist which could be used in wide spectrum of disciplines such as sociology, political sciences, and geology. Social media provides two broad contexts from social scientist perspective; it provides a wide range of data in already well established social science subjects such as political sciences and sociology, and social media sometimes is seen as a fundamental change in underlying assumptions of the social theory [6]. Meanwhile, it is very difficult to connect the social scientific understanding of social to social media data. For example, the concept of conventional friendship hardly applies to the concept of friendship in social media [6].

4.1. Types of Analytics

4.1.1. Descriptive Analytics

Descriptive social media analytics is all about what happened and/or what is happening. Descriptive analytics collects and describes social media data in the form of reports, visualizations, and clusters to understand the actual business problem. Social media user comments analysis, comes under the descriptive analytics category. Comment analysis can be used to understand users' sentiments and identifying the developing trends by clustering themes and topics. Currently, descriptive analytics accounts for the majority of social media analytics site.

4.1.2. Diagnostic Analytics

Diagnostic social media analytics describes about why something happened. While descriptive analytics provides an overview of your social media marketing campaign's performances (posts, mentions, followers, fans, page views, reviews, pins, etc); diagnostic analytics can improve this data into a single view to see what worked in your past campaigns. Diagnostic analytics include inferential statistics, behavioral analytics, correlations & retrospective analysis and outcome being cause and effect analysis of a business issues.

4.2. PROCESS

Social media analytics process has 3 important steps i.e. capture, understand and present, as Figure 2 says.

4.2.1. Capture:

This means define your specific social objective and KPI's. KPI objectives should be measurable.

For example, in the review we took data from Facebook and Instagram of different pages. So, the possible KPIs will be no. of posts of the page, audience of the page, likes, trusted judgement etc.

4.2.2. Understand:

Quickly identify what actions are working from which networks. Provide relevant content to message match your social media companies. If users are not interacting with content on your site, then change it. Ensure all links are tagged including URL shortening links.

4.2.3. Present:

After complete understanding, we need to visualize and present the same. For visualization there are different tools available. In this review we have used Tableau.

4.3. TYPES OF VISUALIZATION TOOLS

4.3.1. Tableau: Tableau is a top data visualization tool. It has many required and unique features. Its data discovery, exploration and visualization application allow user to answer business questions. You can use Tableau's drag and drop interface to visualize any data, explore different views, and can combine multiple databases easily.

- 4.3.2. TrendMiner:** It is an open-source framework for text processing of streaming social media data [7].
- 4.3.3. TwitIE:** TwitIE is also an open-source information extraction pipeline for social media data particularly microblogs. It performs most of the preprocessing tasks like language identification, tokenizing, part-of-speech tagging and normalizing [8].
- 4.3.4. TwitterZombie:** The software is used to gather data for a series of search phrases simultaneously using Amazon's cloud computing platform [9] and supports basic preprocessing on collected data.

4.4. TABLEAU

Tableau is a Business Intelligence tool. It is used to analyze and visualize the data. Users can create and distribute an interactive dashboard, which portray the trends, variations, and density of the data in the form of graphs and charts. Tableau can connect to files, relational and Big Data sources to acquire and process data. This software allows data blending and real-time collaboration. It is used by businesses, academic researchers, and many government organizations for visual data analysis.

After this preprocessing of data, it need to be cleansed. A traditional approach to text data cleaning is to 'pull' data into a spreadsheet or spreadsheet-like table and then reformat the text [10]. Now the real work begins i.e. visualization.

4.5. DATA ANALYTICS

We have discussed above the three-main process of data analytics i.e. *Capture, Understand* and *Present*. Lets go through a little deeper. The capture stage must balance the need for finding information from all quarters (inclusivity) with focusing on sources that are most relevant and authoritative (exclusivity) to assist in more refined understanding [11]. The understand stage is the core of the entire social media analytics process. The success of this stage will have significant impact on the information and metrics that are displayed in the present stage, and thus the success of future decisions or actions that might be taken by a firm. Depending upon the techniques used and the information required, certain analysis may be pre-processed offline while others are computed on-the-fly using data structures optimized for anticipated, ad hoc uses [11]. The last stage in the social

media analytics process is the present stage. The results from different analytics will be summarized, evaluated, and shown to users in an easy to understand format. Different visualization techniques may be used to present useful information. One of the most commonly used interface designs is the visual dashboard, which aggregates and displays information from various sources.

G. F. Khan, suggest that social media analytics is a six steps iterative process (involving both the science and art) of mining the desired business insights from social media data: [12]

Step 1 Identification: Searching and identifying the right source of information for analytical purposes [12].

Step 2 Extraction: Once a reliable and mineable source of social media data are identified, next comes extraction of the data through APIs or manually [12].

Step 3 Cleaning: This step involves removing the unwanted data from the automatically extracted data [12].

Step 4 Analyzing: Next, the clean data is analyzed for business insights. Depending on the layer of social media analytics under consideration and the tools and algorithm employed, the steps and approach to take will greatly vary [12].

Step 5 Visualization: Depending on the type of data, the analysis part will lead to relevant visualizations for effective communication of results [12].

Step 6 Interpretation or Consumption: This step relies on human judgments to interpret valuable knowledge from the visual data. Meaningful interpretation is of particular importance when we are dealing with descriptive analytics that leaves room for different interpretations [12].

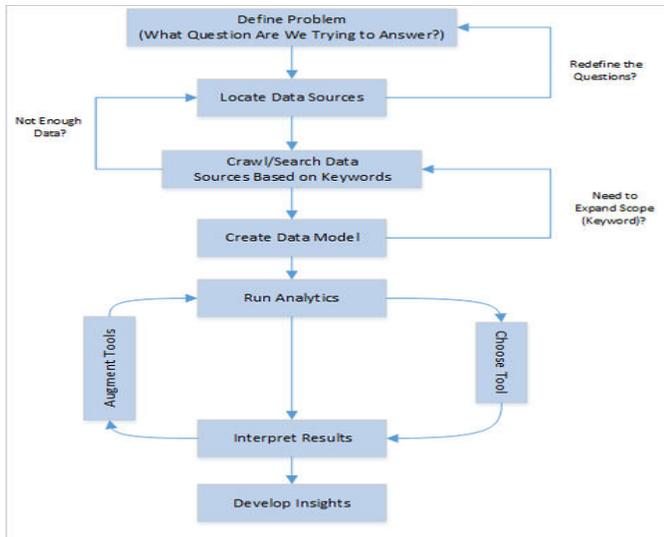


Fig: 2- Data Analytics process by G. F. Khan

4.5.1. DATA CLEANING

Before analyzing this massive dataset, it need to be cleaned. As we know that, social media dataset holds unstructured data. These datasets consist of audio, video, images and text data, geographical data etc. Not only these things but also likes, click views, page rates, comments, shares and replies are present. It may have redundant data or kind of garbage. A traditional approach to text data cleaning is to ‘pull’ data into a spreadsheet or spreadsheet-like table and then reformat the text. For example, Google Refine is a standalone desktop application for data cleaning and transformation to various formats. Transformation expressions are written in proprietary Google Refine Expression Language (GREL) [13].

4.5.2. DATA ANALYSIS

Data analytics (DA) is the process of examining data sets in order to draw conclusions about the information they contain, increasingly with the help of specialized systems and software. Data analytics technologies and techniques are widely used in commercial industries to enable organizations to make more-informed business decisions and by scientists and researchers to verify or disprove scientific models, theories and hypotheses. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains. While analyzing the data we need to use some filters. Now-a-days tableau filters become the leading. Tableau filter is a process to remove certain values or range of values from a result

set. This feature allows both simple situations using field values as well as advanced calculation or context-based filters. There are three types of basic filters available in Tableau such as Filter Dimensions, Filter Measures, Filter Dates.

4.5.3. DATA VISUALIZATION

Visualization is a process of creating interpretable images, diagrams or animations to communicate both abstract and concrete ideas. Visualization helps us better understand data on a graphical platform [18].

For visualization purpose the tableau desktop has been used. By using appropriate dimensions and measures in the appropriate row and column. When we take the instance of visualizing social media data there are many options available for simple to complex visualization.

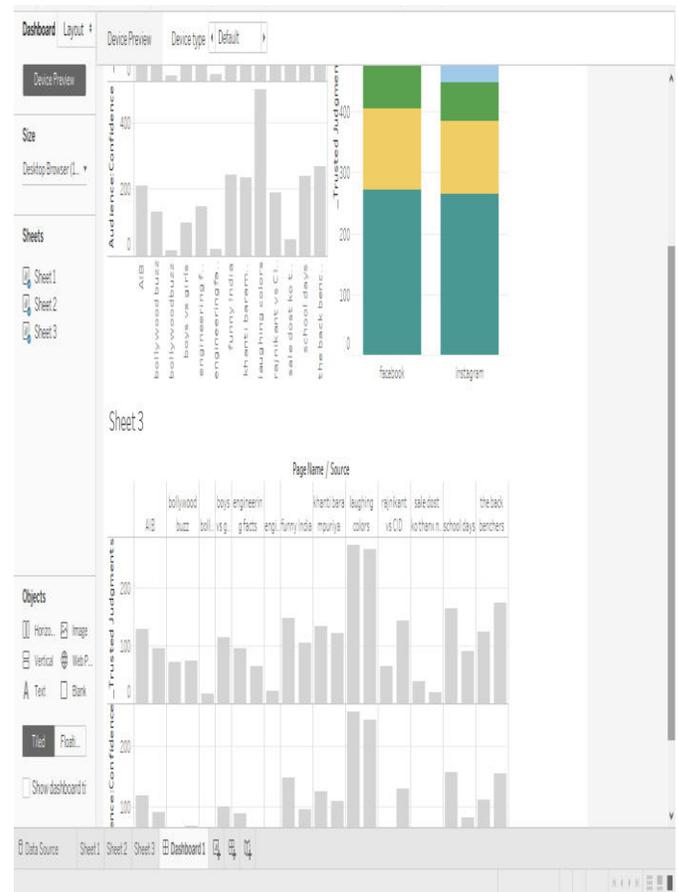


Fig: 3- A Dashboard

To visualize the analyzed data, we need a dashboard. A dashboard is a merged display of many worksheets and related information in a single place. It

is used to compare and monitor a variety of data simultaneously. A good dashboard informs with an informative view. A great dashboard combines high performance and ease of use. So that, anybody can get the data-driven answers to their business questions. Creating a dashboard with Tableau helps to the non-technical users to create interactive, real-time visualizations. By just combining worksheets, i.e. the analyzed data sources, adding some filters, and drill down into specific information. We can share the dashboards, which does not require any scripting or high-level programming.

V. CONCLUSION

Social media data has a lot of knowledge hidden in its huge volumes of noisy, unstructured and dynamic collection of data [4]. This paper is an overview of different phases of in social media analytics. It focuses on various open source tools useful for preprocessing, cleaning, analyzing and visualization process. In this paper, the work done by analyzing various social media pages to know their popularity in the audience, by using visualization.

REFERENCES

- [1] Ralph K. Yeh; Visualization Techniques for Data Mining in Business Context: A Comparative Analysis; University of Texas at Arlington.
- [2] Anita Kumari Singh; Mogalla Shashi: Research Aids for Social Media Analytics: Research Scholar, Department of Computer Science & Systems Engineering, Andhra University, Visakhapatnam, India, Professor, Department of Computer Science & Systems Engineering, Andhra University, Visakhapatnam, India
- [3] Wikipedia contributors. "Sentiment analysis." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 24 Nov. 2017. Web. 9 Dec. 2017.
- [4] Samuel Fosso Wamba, Shahriar Akter, Hyunjin Kang, Mithu Bhattacharya, Mohammed Upal; The Primer of Social Media Analytics; NEOMA Business School, Mont-Saint-Aignan, France, University of Wollongong, Wollongong, Australia, Department of Marketing, George Washington University, Washington D.C., USA, University of Detroit Mercy, Detroit, MI, USA, WebHawks IT, Tokyo, Japan.
- [5] Phillip Brooker, Julie Barnett and Timothy Cribbin: Doing social media analytics
- [6] Industrial & Management Engineering, IIT Kanpur
- [7] Preotiuc-Pietro, Daniel, et al. "Trendminer: An architecture for real time analysis of social media text." (2012)
- [8] Bontcheva, Kalina, et al. "TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text." RANLP. 2013.
- [9] Black, Alan, et al. "Twitter zombie: Architecture for capturing, socially transforming and analyzing the Twittersphere." Proceedings of the 17th ACM international conference on Supporting group work. ACM, 2012.
- [10] Bogdan Batrinca • Philip C. Treleaven: Social media analytics: a survey of techniques, tools and platforms
- [11] Weiguo Fan ¹, Michael D. Gordon ²; Unveiling the power of Social Media Analytics; Virginia Tech ¹, University of Michigan, Ann Arbor ².
- [12] Khan G. F., 2015, seven layers of social media analytics: Mining business insights from social media text, actions, networks, hyperlinks, apps, search engine, and location data, CreateSpace Independent Publishing Platform.
- [13] Bogdan Batrinca, Philip C. Treleaven; Social media analytics: a survey of techniques, tools and platforms.