1.10 SOCIAL MEDIA ANALYTICS CYCLE

- Social media analytics is a six step irrelative process (involving both the science and art) of mining the desired business insights from social media data (Fig. 1.10.1). At the center of the analytics are the desired business objectives that will inform each step of the social media analytics journal.
- Business goals are defined at the initial sage, and the analytics process will continue until the stated business objectives are fully satisfied. To arrive from data to insights, the steps may vary greatly based on the layers of social media mined (and the type of the tool employed).
- The following are the six general steps, at the highest level of abstraction, that involve both the science and art of achieving business insights from social media data.

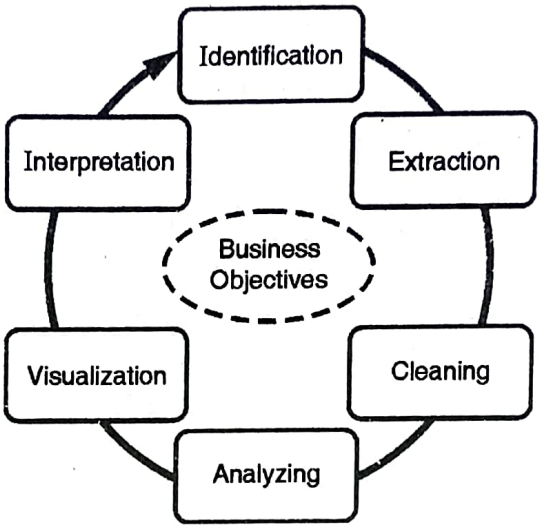

Fig. 1.10.1 : Social media analytics Cycle

Step 1 : Identification

- The identification stage is the art part of social media analytics and is concerned with searching and identifying the right source of information for analytical purposes.
- The numbers and types of users and information (such as text, conversation, and networks) available over social media are huge, diverse, multilingual, and noisy.
- Thus, framing the right question and knowing what data to analyze is extremely crucial in gaining useful business insights. The source and type of data to be analyzed should be aligned with business objectives. Most of the data for analytics will come from your business-owned social media
- platforms, such as your official Twitter account, Facebook fan pages, blogs, and YouTube channel. Some data for analytics, however, will also be harvested from non official social media platforms, such as Google search engine trends data or Twitter search stream data.

- The business objectives that need to be achieved will play an important role in identifying the sources and type of data to be mined.

Step 2 : Extraction

- Once a reliable and minable source of data is identified, next comes the science of extraction stage.
- The type (e.g., text, numerical, or network) and size of data will determine the method and tools suitable for extraction. Small-size numerical information, for example, can be extracted manually (e.g., going through your Facebook fan page and counting likes and copying comments), and a large-scale automated extraction is done through an API (application programming interface). Manual data extraction maybe practical for small-scale data, but it is the API-based extraction tools that will help you get most out of your social
- media platforms. Mostly, the social media analytics tools use API-based data extraction. APIs, in simple words, are sets of routines/protocols that social media service companies (e.g., Twitter and Facebook) have set up that allow users to access small portions of data hosted in their databases.
- The greatest benefit of using APIs is that it allows other entities (e.g., customers, programmers, and other organizations) to build apps, widgets, websites, and other tools based on open social media data. Some data, such as social networks and hyperlink networks, can only be extracted through specialized tools.
- Two important issues to bear in mind here are the privacy and ethical issues related to mining data from social media platforms.
- Privacy advocacy groups have long raised serious concerns regarding large-scale mining of social media data and warned against transforming social spaces into behavioral laboratories.
- The social media privacy issue first came into the spotlight particularly due to the large-scale "Facebook Experiment" carried out in 2012, in
- which Facebook manipulated the news feeds feature of thousands of people to see if emotion contagion occurs without face-to-face interaction (and absence of nonverbal cues) between people in social networks (Kramer, Guillory et al.2014).
- Though the experiment was consistent with Facebook's Data Use Policy(Editorial 2014) and helped promote our understanding of online social behavior, it does, however, raise serious concerns regarding obtaining informed consent from participants and allowing them to opt out.
- The bottom line here is that your data extraction practices should not violate a user's privacy and the data extracted should be handled carefully.
- While all social media platforms have their privacy policies in place, to be on the safe side it is advisable to craft your own social media privacy policy.

- Your policies should explicitly detail social media ownership in terms of both accounts and activities such as individual and page profiles, platform content, posting activity, data handling and extraction, etc.

Step 3 : Cleaning

- This step involves removing the unwanted data from the automatically extracted data. Some data may need a lot of cleaning, and others can go into analysis directly.
- In the case of the text analytics, for example, cleaning, coding, clustering, and filtering may be needed to get rid of irrelevant textual data using natural language processing (NPL).
- Coding and filtering can be performed by machines (i.e., automated) or can be performed manually by humans. For example, Discover Text combines both machine learning and human coding techniques to code, cluster, and classify social media data (Shulman 2014).

Step 4 : Analyzing

- At this stage 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 you take will greatly vary. For example, nodes in a social media network can be clustered and visualized in a variety of ways depending on the algorithm employed.
- The overall objective at this stage is to extract meaningful insights without the data losing its integrity. While most of the analytics tools will follow you through the step-by-step procedure to analyzey our data, having background knowledge and an understanding of the tools and its capabilities is crucial in arriving at the right answers.

Step 5 : Visualization

- In addition to numerical results, most of the seven layers of social media analytics will also result in visual outcomes.
- The science of effective visualization known as visual analytics is becoming an important part of interactive decision making facilitated by solid visualization (Wong and Thomas2004; Kielman and Thomas 2009). Effective visualization is particularly helpful with complex and huge data because it can reveal hidden patterns, relationships, and trends.
- It is the effective visualization of the results that will demonstrate the value of social media data to top management. Depending on the layer of the analytics, the analysis part will result in relevant visualizations for effective communication of results.
- Text analytics, for instance, can result in a word cooccurrence cloud; hyperlink analytics will provide visual hyperlink networks; and location analytics can produce interactive maps.

- Depending on the type of data, different types of visualization are possible, including the following.
  - Network data (with whom) - network data visualizations can show who is connected to whom. For example, a Twitter following-following network chart can show who is following whom. Different types of networks are discussed in a later chapter.
  - Topical data (what) - topical data visualization is mostly focused on what aspect of a phenomenon is under investigation. A text cloud generated from social media comments can show what topics/themes are occurring more frequently in the discussion.
  - Temporal data (when) temporal data visualization slice and dice data with respect to a time horizon and can reveal longitudinal trends, patterns, and relationships hidden in the data. Google trends data, for example, can visually investigate longitudinal search engine trends.
  - Geospatial data (where) geospatial data visualization is used to map and locate data, people, and resources. The chapter on location analytics provides more details on mapping.
  - Other forms of visualizations include trees, hierarchical, multidimensional (chart, graphs, tag clouds), 3-D (dimension), computer simulation, infographics, flows, tables, heat maps, plots, etc.

▶ **Step 6 : Interpretation**

- Interpreting and translating analytics results into a meaningful business problem is the art part of social media analytics.
- This step relies on human judgments to interpret valuable knowledge from the visual data. Meaningful interpretation is particularly important when we are dealing with descriptive analytics that leave room for different interpretations. Having domain knowledge and expertise are crucial in consuming the obtained results correctly.
- Two strategies or approaches used here can be
  (1) producing easily consumable analytical results and
  (2) improving analytics consumption capabilities.
- The first approach requires training data scientists and analysts to produce interactive and easy-to-use visual results. And the second strategy focuses on improving management analytics consumption capabilities.

1.11 CHALLENGES TO SOCIAL MEDIA ANALYTICS

- Social media data is high volume, high velocity, and highly diverse, which, in a sense, is a blessing in terms of the insights it carries; however, analyzing and interpreting it presents several challenges.
- Analyzing unstructured data requires new metrics, tools, and capabilities, particularly for real-time analytics that most businesses do not possess.

Volume And Velocity As A Challenge

- Social media data is large in size and is swiftly generated. Capturing and analyzing millions of records that appear every second is a real challenge. For example, on Twitter, three-hundred forty-two thousand tweets appear every minute, and on Facebook, one million likes are shared every twenty minutes.
- Capturing all this information may not be feasible. Knowing what to focus on is crucial for narrowing down the scope and size of the data. Luckily, sophisticated tools are being developed to handle high-volume and high-velocity data.

Diversity As Challenge

- Social media users and the content they generate are extremely diverse, multilingual, and vary across time and space. Not every tweet, like, or user is worth looking at. A tweet or mention coming from an influential social media user is more important than a tweet from a noninfluential user.
- Due to the noisy and diverse nature of social media data, separating important content from noise is challenging and time consuming.

Unstructuredness As A Challenge

- Unlike the data stored in the corporate databases, which are mostly numbers, social media data is highly unstructured and consists of text, graphics, actions, and relations.
- Short social media text, such as tweets and comments, has dubious grammatical structure, and is laden with abbreviations, acronyms, and emoticons(a symbol or combination of symbols used to convey emotional expressions in text messages), thus representing a great challenge for extracting business intelligence.

1.12 SOCIAL MEDIAANALYTICS TOOLS

- To keep up with the growing need for analyzing the vast amount of data, social media analytical tools are also coming to market at a great pace. Social media analytics tools come in a variety of forms and functionalities.
- Table 1.12.1 lists some example tools with respect to each layer of social media analytics. These tools are briefly discussed in their respective chapters. Aligned with your social media strategy, these tools can be used to measure different layers of social media data.

Table 1.12.1 : Social media analytics tools

Layer of social media	Example of tools
Text	• Discovertext • Lexalytics • Tweet Archivist • Twitonomy • Netlytic • LIWC • Voyant
Actions	• Lithium • Twitonomy • Google Analytics • SocialMediaMineR
Networks	• NodeXL • UCINET • Pajek • Netminer • Flocker • Netlytic • Reach Mentionmapp
Mobile	• Countly • Mixpanel • Google Mobile Analytics
Location	• Google Fusion • Table Location • Google Fusion Table • Tweepsmap • Trendsmap • Followerwonk • Esri Maps • Agos
Hyperlinks	• Webometrics Analyst • VOSON
Research Engines	• Google Trends

Chapter Ends...

□□□