

# A Survey on Social Media Content Analysis

**Sapna Mittal**

Assistant Professor, Computer Science, Post Graduate Department of Computer Science and Applications, R.S.D College,  
Ferozepur, Punjab, India

## ABSTRACT

Social media might even be a platform wherever folks will specific their opinions and concepts and communicate with others from all over the globe. Social media will receive text, audio, video and pic formats. Social media promoting might even be the simplest way that created it achievable for corporations to introduce their merchandise and services at intervals web, and reach a community and customers that they will not reach with ancient promoting channels. Content analysis on social media will facilitate corporations to form their marketing ways that corporations will verify the recognition of their merchandise and sentiment analysis are usually accustomed study the public response regarding the company merchandise.

**Keywords** : Social Media, Social Media Content Analysis, Web-based social networking

## I. INTRODUCTION

The Internet has attracted the eye of study communities specially, the various role of analyzing social media and networks to advance our understanding of data sharing, communication, opinion formation, and dissemination has been recognized. Still, rigorous, quantitative studies on social media content, notably on electronic commerce and knowledge management, keep scarce. The foremost extended barrier to social media usage is that the dearth of a flexible methodology for choosing, collecting, processing, and analyzing discourse knowledge obtained from social media sites. However, many code corporations have developed proprietary text mining systems for information image, and researchers have developed skilled systems for sentiment analysis. Nevertheless, social media content is wide accessible, up-to-date, and accessible in electronic format. Therefore, a scientific approach is essential, as a result of it helps electronic commerce researchers, organizations, and governments understand the commonality in numerous on-line text data that appear in social media. Misinterpreting

the info obtained from social media, researchers can gain valuable insights into the beliefs, values, attitudes, and perceptions of social media users with connection the utility of user generated content and trust formation. Consequently, such knowledge can facilitate marketers monitor the perceptions of individuals relating to social networks and aid organizations in strategic planning to deal with the gap between the availability of user generated raw text and thus the discourse knowledge of mass data. This study introduces a grounded theory approach to research social media content to spot the underlying issue structure of the collected knowledge and to interpret the structure in regard to the study objective. Strauss and Corbin [1998] made public the grounded theory approach as a quest methodology that employs a scientific set of procedures to develop associate degree “inductively derived” grounded theory a couple of specific development. This grounded theory can likewise be accustomed explore ideas and develop themes supported qualitative data.

## II. LITERATURE REVIEW

**2.1 Social Media** : Online networking contains internet-based applications that are produced in light of the ideological and innovative establishments of Web 2.0. Online networking empowers the creation and trade of client produced content [Hinchcliffe, 2008; Karimov et al., 2011; Turban et al., 2015]. Utilizing Internet-and online innovations, web-based social networking change communicate media monologs (i.e., one-to-many) into online networking exchanges (i.e., many-to-many). Through social media, clients can transfer photographs, recordings, music, pictures, and messages to share thoughts, emotions, assessments, and encounters with different individuals [Lai and Turban, 2008; Turban et al., 2015]. In creating nations, for example, China and India, social media has experienced marvelous development [Lai and To, 2012; To et al., 2014; Srivastava and Pandey, 2013]. Specifically, long range informal communication destinations, online discussions, texting administrations, and versatile shrewd stages have developed exponentially, bringing about the across the board utilization of online networking. In such manner, web-based social networking has turned into a intense power of democratization. Web-based social networking empowered correspondence and cooperation among people at a huge scale without land, time, and framework imperatives [Hinchcliffe, 2008; Lai and Turban, 2008]. The individual components of online networking groups incite abnormal amounts of trust. Such trust brings about the impression of the gotten data's dependability [Karimov et al., 2001]. Trust and data trade are basic segments of basic leadership.

**2.2 Social Media Content Analysis** : The apace increasing quantity of social media info and shopper views on a product or service, which can be either positive or negative, incorporates a considerable result on a company. Thus, researchers have developed sophisticated tools for topic modeling and document agglomeration [Banerjee & Basu, 2007; Becker et al., 2009; Blei et al., 2003; Ramage et al., 2011; Ma et al.,

2013], still as text mining tools [Aggarwal & Wang, 2011; Hu & Liu, 2012; Morinaga et al., 2002; Ye et al., 2009]. The unattended learning of latent topics is beneficial for various on-line applications, like organizing documents consistent with topic-based agglomeration, and data filtering supported user preferences [Banerjee & Basu, 2007]. The subject model utilizes the theorem model for text document assortment [Blei et al., 2003]. It mechanically learns a group of thematic topics from collected documents and so assigns variety of these topics to every collected document. The subject model may be thought of as a probabilistic version of latent semantic analysis [Deerwester et al., 1990; Newman & Block, 2006].

## III. Methodology

### Phase I: Define Goal

The most essential initiate any project is that the definition of the goal and scope of the study. Several specific criteria on the dataset is known and delineated . A content analysis can become associate infinite study if the objectives of the investigation square measure one.

### Phase II: Collecting data

Having outlined a goal and a clear scope, the investigator can then confirm the factors, together with the sources and additionally the variability of sites to be downloaded. The sources of sites have to be compelled to be identified supported the scope of the study.

### Phase 3: Data Transformation

In the past, content analysis was mostly conducted manually, with investigators coding text by classification, categorization, and subjective interpretation. With the advancement of lexical code, linguistics code, and applied mathematics tools, investigators can objectively interpret qualitative knowledge from a wider perspective by distinctive underlying key attributes, factors etc.

Phase 4: Results

In the final section of the methodology, the investigator ensures that the results obtained in section 3 square measure explicable, specifically with relevancy the goal and scope made public in section one. The investigator summarizes the findings, identifies the group action and smart implications of the findings to assist in decision-making, and acknowledges the restrictions of the study.

#### IV. CONCLUSION

As the net will increase in size, mode, and variety, exploring online page, notably social media, and transforming such content into ideas became a challenge to electronic commerce researchers, business practitioners, and policymakers. During this paper, we tend to introduced a scientific methodology to convert text files from social media to ideas that area unit repeatable, simply explainable, and visual with an idea map. We've likewise established many criteria to spot sources, minimum sample size (i.e., total range of webpages), the sample size of each class of sources, and also the range of key variables (or attributes). The grounded theory approach indicates that other than collection relevant information for analysis, we tend to additionally got to allow ideas and themes to emerge from the bottom up. User-generated content, like viva-voce, can be systematically monitored to know the beliefs, values, attitudes, perceptions, intentions, and behaviors of users. The projected methodology, whether or not quantitative or qualitative in nature, advantages business practitioners, considering that cost, time, and human errors area unit unbroken to a minimum throughout processing and analysis.

#### V. REFERENCES

[1]. Averya, E., R. Lariscyb, and K.D. Sweetserb, "Social Media and Shared- or Divergent-uses? A Coorientation Analysis of Public Relations Practitioners and Journalists," *International Journal of Strategic Communication*, Vol. 4, No.3:189-205, 2010.

[2]. Bai, X., "Predicting Consumer Sentiments from Online Text," *Decision Support Systems*, Vol. 50, No. 4:732-742, 2011.

[3]. Banerjee, B., and S. Basu, "Topic Models over Text Streams: A Study of Batch and Online Unsupervised Learning," *Proceedings of the SIAM International Conference on Data Mining (SDM)*, 2007.

[4]. Becker, H., M. Naaman, and L. Gravano, "Event Identification in Social Media," *Proceedings of the ACM SIGMOD Workshop on the Web and Databases (WebDB'09)*, 2009.

[5]. Beerli, A. and J.D. Martin, "Factors Influencing Destination Image," *Annals of Tourism Research*, Vol. 31, No. 3:657-681, 2004.

[6]. Blei, D.M., A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning*, Vol. 3: 993-1022,2003.

[7]. Bondi, M. and M. Scott, *Keyness in Texts*, Amsterdam: John Benjamins B.V., 2010.

[8]. CenStatd, *Hong Kong Annual Digest of Statistics - 2013 Edition*, Hong Kong: Hong Kong Census and Statistics Department, 2013.

[9]. Chiu, C.M., M.H. Hsu, and E.T.G. Wang, "Understanding Knowledge Sharing in Virtual Communities: An Integration of Social Capital and Social Cognitive Theories," *Decision Support Systems*, Vol. 42, No. 3:1872-1888, 2006.

[10]. Cliff, N, "The Eigenvalues-Greater-than-One Rule and the Reliability of Components," *Psychological Bulletin*, Vol. 103, No. 2:276-279, 1998.