# Social Media Analytics and Intelligence

A. Siddhartha Rao
UG Scholar
Dept. of Electrical Engineering
SVITS, Indore, India

Kunal Jagdale
UG Scholar
Dept. of Electrical Engineering
SVITS, Indore, India

## ABSTRACT

Recent advances in the internet have helped social media to influence many areas of business, marketing, weather forecasting, communication etc. Social media Analytics has become a vital part of the data analytics as the vast population is involved in social media and immeasurable insights can be taken from their activities in the social media websites. In this paper, we summarise different techniques used to analyze social media activities like tweets, blogs, etc., and to present the pros and cons of each.

## Keywords

Keywords are your own designated keywords which can be used for easy location of the manuscript using any search engines.

## 1. INTRODUCTION

Data is ubiquitous and has become a vital force in today's world, be it business, defense, weather forecasting, marketing, etc., one can apply analytical tools to prescience the trends to make important decisions for the benefit of the company and the society [1].

Over last decade, the internet has created many new channels and opportunities for companies advertise products and transact business. The introduction of Web 2.0 with its emphasis on information sharing and user collaboration has been altering this landscape [2]. Increasingly, the main focus of discussion of all aspects of a company's product portfolio is moving from their respective websites to collaborative sites, blogs and forums collectively known as social media.

### What is social media analytics?

Social media analytics is an act of collecting and interpreting the data posted on social media in order to conduct market research, learn about a target audience, collect feedback that drives strategic business decisions, etc [3]. With the growing number of population on social media, it's important to know what people talk about media whether it's for business, scientific or sociological reasons.

It is concerned with developing informatics tools and frameworks to collect, monitor, analyze and visualize social media data, usually driven by specific requirements from a target application [4].

### Why social media analytics?

Social media has become a critical part of the information ecosystem and as social media platforms and applications gain widespread adoption with unparalleled reach to users, consumers, voters, businesses and nonprofit organizations alike, interest in social media from every field has been escalated from both application and research perspectives [5].
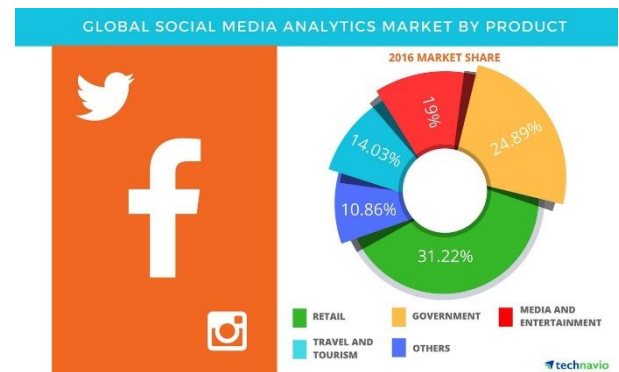


**Fig 1**

As the population on the social networking sites is skyrocketing, many insights can be taken from social media. Currently, there are 2 billion internet users on an approx who are using social networks and the graph is growing as mobile device usage and mobile social networks increasingly gain traction [6]. There are about –

• 500 Million Tweets sent each day

• 4 Million Hours of content uploaded to YouTube every day

• 3.6 Billion Instagram Likes each day

• 4.3 Billion Facebook messages posted daily

• 5.75 Billion Facebook likes every day

• 40 Million Tweets shared each day

• 6 Billion daily Google Searches

All these content comprise more than 50% of total internet data but only a handful of data is analyzed due to lack of techniques and tools.

With the rapid increase in the economy and online sale through e-commerce sites social media is gaining a significant role in today's business world [7]. Every brand is focusing on social media for the betterment of their business and due to this a gradual accretion of data can be observed.

But the question arises how to evaluate such a significant amount of data?

In this paper, we have reviewed different methods to deal with the data generated by social media in every aspect of marketing, weather forecasting, communication and mentioned pros and cons of each.

## 2. MARKETING

As social media and e-commerce become increasingly embroil in our lives, the opportunities for them to interact with and bolster each other are innumerable, considering that the average person spends around an hour and 40 minutes browsing social media every day, and the number of internet shoppers in the US will reach 217 million by 2018 [8].

Back in the 90's, a business presence would be signified by advertisements in the paper and a physical storefront. Now, in the digital age, business reputations live and die by their social media standing. Right now, social media is used by brands as a way to advertise, increase their online presence, and deliver high-quality customer service [9]. In the coming years, we can expect those trends to continue, as some new ones emerge. Due to this increase in the e-commerce, social media analytics has a crucial role in business intelligence.

From a marketing perspective, blogs are a vital form of social media because they provide insights into previously inaccessible information such as opinions customer reviews. A typical blog or micro-blog has one author and consists of multiple entries or posts [10]. A blog can be imagined in a three-dimensional space defined by RELEVANCE, SENTIMENT, AUTHORITY. All three dimensions are equally important and are needed to be considered in a unified view in order to provide marketing insights.

In order to find the relevance, one has to filter the whole blogosphere from millions to thousands of blogs which are relevant to the topic of interest, a topic can be a specific product and to consider each and every blog and to find what impact it's having on the marketability of the product. If the interest is in the specific product it is straightforward to identify relevant blogs containing references to the product [11]. A company can take valuable insights from that particular blogs and implement particular measures to make that product better according to the need of the people.

Another dimension defined is the sentiment, sentiment detection is the crucial part as it is used to identify blog posts that may require swift marketing actions. Every expression of sentiment is domain specific which is the main challenge as they have to be monitored regularly. Here sentiment classifiers come into play as they can rapidly adapt to new domains without requiring a large number of manually labeled training examples of positive and negative sentiments [12]. An approach is used known as active dual supervision in which a word is labeled as negative or positive and the algorithm detects it accordingly.

From a marketing perspective, it's important to identify influential bloggers who are most responsible for the spread of information in the blogosphere, since any negative sentiments they express can reach far and wide. The authority of a blog can be characterized based on the number and authority of other blogs that link to it, using page rank algorithm [13]. For the purpose of social media marketing, we are interested in bloggers who influence the thinking and consider the content blogged by others.
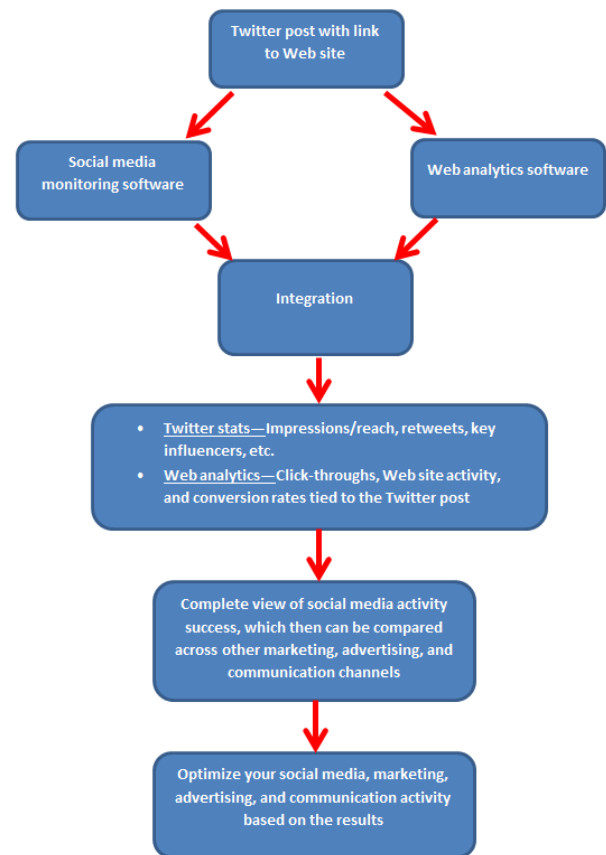


**Fig 2**

## 3. POLITICAL COMMUNICATION

Due to the rapid growth of social media, particularly on Twitter and Facebook, social media is becoming a vital tool in political context – both by citizens and political foundations. From the perspective of political institutions, it has become important to actively participate in political communication through social media, especially during election campaigns [14]. Social media, therefore, represents the ideal platform and information base to gauge public opinion on policies and political positions as well as build community support for candidates running for public offices.

It is proved from the previous studies that from the perspective of political institutions and government agencies, there is a need to gather monitor, analyze, summarize and visualize politically relevant data from social media to improve communication with the citizens and to identify emerging topics and different issues faced by the people [15]. There are three main types of social media namely: -

### 3.1 Microblogging

Microblogging is a combination of blogging and instant messaging that allows users to create short messages to be posted and shared with an audience online. Currently, Congress members consider Twitter rather a vehicle for self-promotion as they are primarily links to news articles about themselves and to their blog posts, and to report on their daily activity [16]. Twitter is widely used for dissemination of politically relevant information and that the mere number of party mentions accurately reflects the election result suggesting that microblogging messages on twitter seem to validly mirror the political landscape [17].

## 3.2 Social Network Sites (SNS)

Facebook has become an increasingly popular method for studying socio-political issues. Such public contributed contents, mainly available as wall posts and corresponding comments on Facebook pages or Facebook groups, influence people to express their opinions and sentiments on a given topic, persons, simultaneously allowing social and political scientists to conduct analyses of political discourse [18]. Facebook is a legitimate location for discussion of political issues and to some extent, the discussion appears to have succeeded in overcoming polarization of online discussion.

## 3.3 Weblogs

Weblogs play a most influential part in an election campaign since political blogs contain information about news articles from mass media, introduction to other blogs postings, and criticism about mass media coverage of political affairs [19]. In traditional mass media and older genres securitization was a problem faced by the political leaders as terrorism has become a major problem but political blogs might provide spaces for securitization discourse that might not be found in previous practices. After the 2008 US presidential election, Wattal et al. (2010) investigated the contingent impact of political blogs on the campaign process [20]. Their results showed that the blogosphere can influence the campaign process and the election outcomes.

## 4. ANALYSIS PURPOSE FOR POLITICAL COMMUNICATION

From the perspective of political party's, there are two major purposes of conducting social media analytics namely: -

## 4.1 Reputation & Impression Management

Many political institutions and parties have the need to be updated about their own reputation and impression in public social media. Following aspects are relevant for the above purpose [21]. First, political parties are interested in how people are talking about them, particularly regarding specific characteristics such as trustworthiness, determination, success etc. Second, they are also interested in emerging topics that have the potential to trigger a crisis or scandal that can harm the reputation. Thirdly, Political parties strike for measure the degree of influence they have in social networks [22].

## 4.2 General Monitoring

Besides management of reputation and impression, political party's might employ social media analytics to monitor the social web in an explorative way. For example, they might be interested in knowing about what kind of political topics or issues are discussed and how these discussions take place in social media [23]. In addition, early detection of upcoming "hot" topics or issues might enable political parties to react timely to such trends. Generally, trends can also be viewed as a reflection of societal concerns or even as a consensus of collective decision making [24].
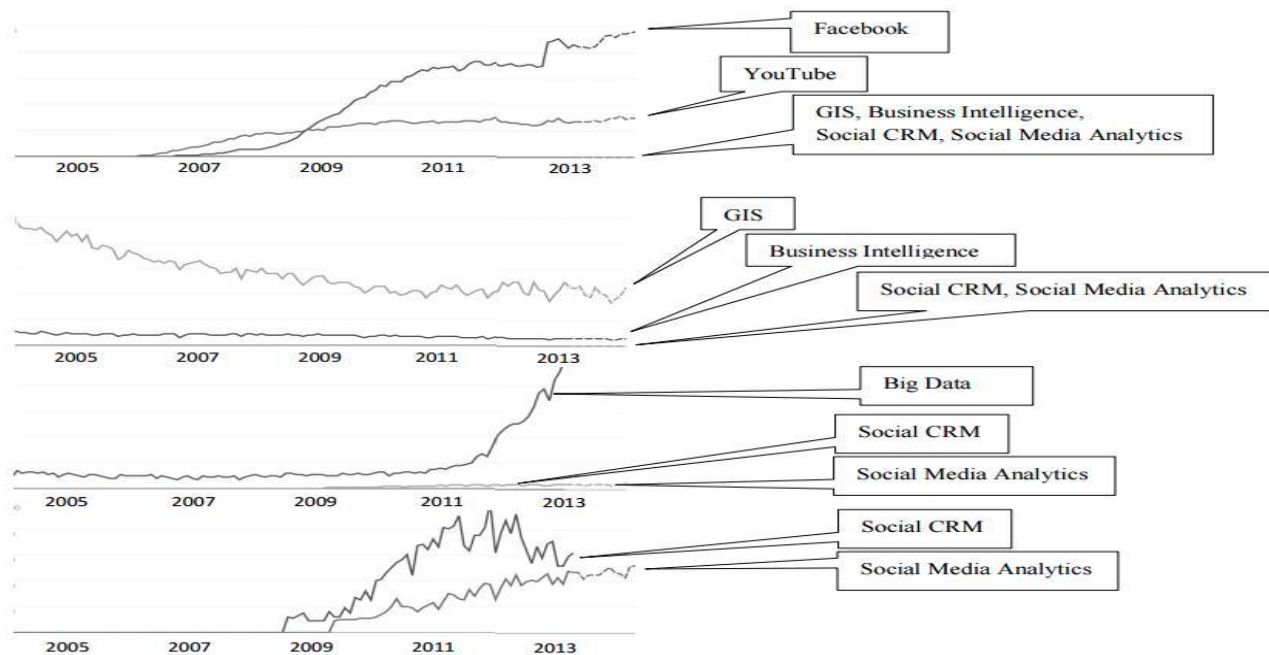


**Fig 3: CRM vs ERP, CRM and ERP**

## 5. ANALYSIS APPROACHES

Besides analysis purpose, another dimension of the framework represents the different analysis approaches. Three common approaches are: -

## 5.1 Topic/Issue/Trend Based Approach

For political parties, it is important to identify and monitor political topics, particularly those that might have direct or indirect contact with themselves as issue contain conflict potential and may evolve into a crisis [25]. In order to identify topics, content analysis or text mining techniques are used to take out relevant data from the social network. Content analysis is a research technique for making relevant and valid inferences from texts to the contents of their use. For massive amounts of social media data, automated quantitative methods of content analysis or text mining are necessary [26].

## 5.2 Opinion/Sentiment Based Approach

Due to the acclivity in the usage of social media, people are enabled to express their views, opinions or emotions on almost anything in forums, blogs, and on SNS more than ever before. Opinions are important as there are the building stones of decision making, one has to consider every opinion while taking a decision which implies particularly to political communication [27]. Sentiment analysis or opinion mining has emerged as a distinct method for the study of people's opinions in terms of views, attitudes, appraisals, and emotions towards entities, events and their attributes in a more thorough way.

## 6. SOCIAL MEDIA ANALYTICS TECHNIQUES

### 6.1 Spatiotemporal Social Media Data Analytics

Twitter has more than 200 million Tweets each day, which has an enormous amount of irrelevant data under which the relevant messages for situational awareness are usually buried [28].

Here comes the spatiotemporal social media analytics approach for abnormal topic detection and event scrutinization. In this approach, Latent Dirichlet Allocation (LDA) which extracts and probabilistically ranks major topics contained in textual parts of the social media data. The ranks give volume based importance [29].

As social media platforms are moving towards location-based networks researchers have proposed various approaches to analyze spatiotemporal data. MacEachren et al. demonstrate a visualization system that denotes the message density of actual or textually inferred Twitter message locations. Their work also has shown that social media can be a potential source for crisis management [30]. Special LBSN for certain domain like bailey and every trail have a stronger focus on the sharing and tracking of user locations. There are much-related works for special nontemporal document collections.

### 6.2 Sparse Multivariate Regression

In the blogosphere, the main obstacle is to identify the most influential participants, here sparse multivariate regression is used.

Multivariate regression is a generalization of the classical regression model of regressing a single response on p predictors to regressing q > 1 responses on p predictors [31]. Applications of this general model arise in chemometrics, econometrics, psychometrics, and other quantitative disciplines where one predicts multiple responses with a single set of prediction variables.

Sparse multivariate regression algorithm which simultaneously performs dimensionality reduction and parameter estimation. The algorithm simultaneously performs dimension reduction and coefficient estimation and automatically estimates the number of latent factors from the data.

Granger Causality is also used for ranking the key influencers as it identifies the influencer by scrutinizing its contents and seeing that their collective past content is predictive of the future content of the blogger [32].

### 6.3 Stream Processing

Increasingly, analytics applications that consume real-time social media, financial 'ticker' and sensor networks data need to process high bundle of temporal data with low latency. These applications require support for online analysis of rapidly changing data streams [33]. However, traditional database management systems (DBMSs) have no pre-defined notion of time and cannot handle data online in real time. This has led to the development of Data Stream Management Systems (DSMSs) (Hebrail 2008)—processing in main memory without storing the data on disk—that handle transient data streams online and process continuous queries on these data streams. Example commercial systems include Oracle CEP engine, Stream Base, and Microsoft's StreamInsight [34].

### 6.4 Scientific Programming Tools for Social Media Analysis

There are numerous computational tools available in the market for analytics but few tools are used on a regular basis because of their accuracy and flexibility, some of them are MATLAB, Python, SQL, Hadoop.

Data processing and data modeling, e.g., regression analysis, are straightforward using MATLAB, which provides time-series analysis, GUI and array-based statistics [35]. MATLAB is significantly faster than the traditional programming languages and can be used for a wide range of applications. Moreover, the exhaustive built-in plotting functions make it a complex analytics toolkit.

Python can be used for (natural) language detection, title and content extraction, query matching and when used in conjunction with a module such as sci-kit-learn [36], it can be trained to perform sentiment analysis, e.g., using a Naïve Bayes classifier.

Another tool used is Hadoop, the name is derived from a cute toy elephant but Hadoop is all but a soft toy. Hadoop is an open source project that offers a new way to store and process big data. The software framework is written in Java for distributed storage and distributed processing of very large datasets on computer clusters built from commodity hardware [37]. Hadoop is a highly scalable storage platform because it can store and distribute very large data sets across hundreds of inexpensive servers that operate in parallel. A key advantage of using Hadoop is its fault tolerance. When data is sent to an individual node, that data is also replicated to other nodes in the cluster, which means that in the event of failure, there is another copy available for use [38].

## 7. CONCLUSION & FUTURE SCOPE

Social media analytics and intelligence research continue to flourish and the need to analyze the data generated by social networking sites is in demand. The initial e-commerce phase of the internet has given rise to a new analytics to better understand consumer preferences and hence create targeted marketing campaigns. Social media marketing has created even bigger challenges than the first wave of internet marketing particularly the need to characterize authority and sentiments relative to a specific topic in a network.

Social media has also become an important political communication channel. Political activities might gain more transparency and citizens might be more involved in the process of decision making through social media. The main obstacle faced by the political organizations is the lack of knowledge of politicians about current topics and how to handle them in social media. This can be prevented by organizing various events and seminars internally to introduce politicians to this growing field.

There is an immense scope in analyzing the social media data and the only thing we have to keep in mind is "we have to start observing things in a different manner".

In this growing era of technology, surplus amount data is flowing through our surrounding and the main obstacle is to analyze this data for future decision making. Analyzing data and taking relevant insights from it has become a vital task for every organization and in every field. The advent of social media and the gigantic part of our population in it has created many platforms for the organizations like advertising companies, political parties to market their products and for the betterment of their election campaigns. With the increasing relevance of and the need of analyzing the political discussions on different social media, this field has promising future aspects. A more and more people engage in social media a tremendous amount of data will be generated day by day and analyzing this will be the building stone of decision making.

# 8. REFERENCES

[1] Sharp, Seena. Competitive intelligence advantage: how to minimize risk, avoid surprises, and grow your business in a changing world. John Wiley & Sons, 2009.

[2] Haklay, Muki, Alex Singleton, and Chris Parker. "Web mapping 2.0: The neogeography of the GeoWeb." Geography Compass 2, no. 6 (2008): 2011-2039.

[3] Ordanini, Andrea, Lucia Miceli, Marta Pizzetti, and Arun Parasuraman. "Crowd-funding: transforming customers into investors through innovative service platforms." Journal of service management 22, no. 4 (2011): 443-470.

[4] Stieglitz, Stefan, and Linh Dang-Xuan. "Social media and political communication: a social media analytics framework." Social Network Analysis and Mining 3, no. 4 (2013): 1277-1291.

[5] Hagel III, John, John Seely Brown, and Lang Davison. The power of pull: How small moves, smartly made, can set big things in motion. Basic Books (AZ), 2012.

[6] Rheingold, Howard. Smart mobs: The next social revolution. Basic books, 2007.

[7] Kaplan, Andreas M., and Michael Haenlein. "Users of the world, unite! The challenges and opportunities of Social Media." Business horizons 53, no. 1 (2010): 59-68.

[8] Geronimo, Michelle. "Online Browsing: Can, Should, and May Companies Combine Online and Offline Data to Learn about You." Hastings Sci. & Tech. LJ 9 (2017): 211.

[9] McGucken, Elliot. "22nets: method, system, and apparatus for building content and talent marketplaces and archives based on a social network." U.S. Patent Application 11/523,309, filed September 19, 2006.

[10] Chen, Chen. "Diffusion Vs. Censorship: An Exploration Into the Use of Social Media by Domestic and International Non-governmental Organizations in Mainland China." PhD diss., University of Florida, 2010.

[11] Melville, Prem, Wojciech Gryc, and Richard D. Lawrence. "Sentiment analysis of blogs by combining lexical knowledge with text classification." In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1275-1284. ACM, 2009.

[12] Liu, Bing. "Sentiment analysis and opinion mining." Synthesis lectures on human language technologies 5, no. 1 (2012): 1-167.

[13] Parmelee, John H., and Shannon L. Bichard. Politics and the Twitter revolution: How tweets influence the relationship between political leaders and the public. Lexington Books, 2011.

[14] Stieglitz, Stefan, and Linh Dang-Xuan. "Social media and political communication: a social media analytics framework." Social Network Analysis and Mining 3, no. 4 (2013): 1277-1291.

[15] Engel-Cox, Jill A., and Raymond M. Hoff. "Science–policy data compact: use of environmental monitoring data for air quality policy." Environmental Science & Policy 8, no. 2 (2005): 115-131.

[16] Golbeck, Jennifer, Justin M. Grimes, and Anthony Rogers. "Twitter use by the US Congress." Journal of the Association for Information Science and Technology 61, no. 8 (2010): 1612-1621.

[17] Stieglitz, Stefan, and Linh Dang-Xuan. "Political communication and influence through microblogging-- An empirical analysis of sentiment in Twitter messages and retweet behavior." In System Science (HICSS), 2012 45th Hawaii International Conference on, pp. 3500-3509. IEEE, 2012.

[18] Larsen, Henrikke Hovda, Johanna Margareta Forsberg, Sigrid Viken Hemstad, Raghava Rao Mukkamala, Abid Hussain, and Ravi Vatrapu. "TV ratings vs. social media engagement: Big social data analytics of the Scandinavian TV talk show Skavlan." In Big Data (Big Data), 2016 IEEE International Conference on, pp. 3849-3858. IEEE, 2016.

[19] Stieglitz, Stefan, and Linh Dang-Xuan. "Social media and political communication: a social media analytics framework." Social Network Analysis and Mining 3, no. 4 (2013): 1277-1291.

[20] Wattal, Sunil, David Schuff, Munir Mandviwalla, and Christine B. Williams. "Web 2.0 and politics: the 2008 US presidential election and an e-politics research agenda." MIS quarterly (2010): 669-688.

[21] Stieglitz, Stefan, and Linh Dang-Xuan. "Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior." Journal of Management Information Systems 29, no. 4 (2013): 217-248.

[22] Star, Susan Leigh, and Karen Ruhleder. "Steps toward an ecology of infrastructure: Design and access for large information spaces." Information systems research 7, no. 1 (1996): 111-134.

[23] Stieglitz, Stefan, and Linh Dang-Xuan. "Social media and political communication: a social media analytics framework." Social Network Analysis and Mining 3, no. 4 (2013): 1277-1291.

[24] Koppenjan, Johannes Franciscus Maria, and Erik-Hans Klijn. Managing uncertainties in networks: a network approach to problem solving and decision making. Psychology Press, 2004.

[25] Kitschelt, Herbert. "Linkages between citizens and

politicians in democratic polities." Comparative political studies 33, no. 6-7 (2000): 845-879.

[26] Stieglitz, Stefan, and Linh Dang-Xuan. "Social media and political communication: a social media analytics framework." Social Network Analysis and Mining 3, no. 4 (2013): 1277-1291.

[27] Eisenhardt, Kathleen M. "Making fast strategic decisions in high-velocity environments." Academy of Management journal 32, no. 3 (1989): 543-576.

[28] Chae, Junghoon, Dennis Thom, Harald Bosch, Yun Jang, Ross Maciejewski, David S. Ebert, and Thomas Ertl. "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition." In Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on, pp. 143-152. IEEE, 2012.

[29] Li, Liyun. "Mining Generic Patterns and Communities from Heterogeneous Network Traffic." PhD diss., Polytechnic Institute of New York University, 2013.

[30] Chae, Junghoon, Dennis Thom, Harald Bosch, Yun Jang, Ross Maciejewski, David S. Ebert, and Thomas Ertl. "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition." In Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on, pp. 143-152. IEEE, 2012.

[31] Yuan, Ming, Ali Ekici, Zhaosong Lu, and Renato Monteiro. "Dimension reduction and coefficient estimation in multivariate linear regression." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69, no. 3 (2007): 329-346.

[32] Niles, Travona L. "Academic and Psychological Aspects of Collegiate Cheerleading Participation." PhD diss., Trevecca Nazarene University, 2016.

[33] Chen, Yixin, Guozhu Dong, Jiawei Han, Benjamin W. Wah, and Jianyong Wang. "Multi-dimensional regression analysis of time-series data streams." In Proceedings of the 28th international conference on Very Large Data Bases, pp. 323-334. VLDB Endowment, 2002.

[34] Zaharia, Matei, Tathagata Das, Haoyuan Li, Timothy Hunter, Scott Shenker, and Ion Stoica. "Discretized streams: Fault-tolerant streaming computation at scale." In Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles, pp. 423-438. ACM, 2013.

[35] McKinney, Wes. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc.", 2012.

[36] Das, Prajit Kumar. "Context-dependent privacy and security management on mobile devices." Ph. D. Dissertation (2017).

[37] Shafer, Jeffrey, Scott Rixner, and Alan L. Cox. "The hadoop distributed filesystem: Balancing portability and performance." In Performance Analysis of Systems & Software (ISPASS), 2010 IEEE International Symposium on, pp. 122-133. IEEE, 2010.

[38] Shvachko, Konstantin, Hairong Kuang, Sanjay Radia, and Robert Chansler. "The hadoop distributed file system." In Mass storage systems and technologies (MSST), 2010 IEEE 26th symposium on, pp. 1-10. IEEE, 2010.