

Analysis of Systematic Data Mining Approaches for Achieving Competitive Advantage by Monitoring Social Media

MMF Naja¹, MII Mohamed²

¹# Department of Information and Communication Technology, South Eastern University of Sri Lanka, University Park, Oluvil, Sri Lanka

²Business Intelligence Division, IFS R & D International Private Limited, Colombo - 06, Sri Lanka

mmfnaja@seu.ac.lk

Abstract— This research paper mainly focuses at clarifying how and why social media monitoring is carried out to achieve competitive advantage. A systematic literature review was used to find current insights on the methods used, and so illustrate not only the benefits but also the difficulties that may arise during the monitoring process. Companies use social media to improve their business. More and more companies nowadays use social media tools such as Facebook, Twitter, linked in, you tube, Google+ etc. not only for the purpose of service providing, but also to get in contact with customers.. Resulting this, a large amount of user-generated content is freely available on social media sites which is being utilized by business stakeholders. To increase competitive advantage and critically evaluate the competitive environment of businesses, companies are in need of monitoring and analysing not only the customer-generated content in the company's social media sites, but also the information on their competitors' social media sites as well as the customers. In an effort to aid firms understand how to perform a social media monitoring for achieving competitive advantage and convert data from social media into knowledge that will aid decision making. This review paper also describes an in-detail how text mining provides a way to analyse unstructured text content on social media sites. The results shows the value of social media and the advantages of text mining as an effective way to gain business value from social media data available in plenty. Moreover over, some recommendations are also given in this paper that aid business firms to develop competitive strategy.

Keywords— Sentiment Analysis, Text Mining, Social Network Sites, Opinion Mining, Big Data Analysis

I. INTRODUCTION

Social media have extremely changed our lives and how we interact with one another and our surrounding world. Recent survey shows that most number of people use social media applications for various types of reasons

such as connecting with new friends, contacting old friends, exchanging information, and for making friends. As a result, many business firms are adopting social media to get into this growing trend in order to achieve business values such as marketing management, competitor benchmarking, influencer identification and outreach, market research and reputation management. Social media monitoring supports typical activities such as customer care, sales, branding (advertising, marketing, and content delivery), and support, product development and innovation. Social media tools that are widely used nowadays have generated a collection of textual data, which contain hidden knowledge about businesses to leverage for attaining a competitive edge. In particular, firms can go deep into the variety of amount of social media data to discover and detect new knowledge (e.g., brand popularity) and interesting patterns, get to know what their competitors are doing and how the industries are changing their trends, and use the findings and improved understanding, to achieve competitive advantage against their competitors (Xing & Zhan, 2015).

II. CRITICAL EVALUATION OF LITERATURE

A. What makes social media?

Social media platforms commonly characterized by the profile-based communication and networking of people via the Internet. The focus is addition to disclosing the own personality on a user profile especially the collaborative exchange within a larger community. Under the core aspect of the interest-governed networking People thus find, for example, together in thematically appropriate groups and get the opportunity to talk about specific issues. An important feature of this type of communication within social media networks is the joint design of content through the use of text as well as the aid of image, audio and video content. Moreover, the aspect of networking implies the construction of secondary contacts on primary contacts that would come hardly possible without social media platforms.

B. Why should business firms care about it?

Within the last one decade, social media has taken the full control of the world. Facebook being the king of social networking sites, it holds more than 750 million active people using it. Even though, the majority of the business firms did not immediately join the social media revolution. Instead, they were focusing on some other modes that aided them in developing and improving their businesses. But when the first group of companies joined, some of the rest of the people anticipated the importance of it. Many others, however, remained shutting, mainly because of a lack of knowledge and an unclear understanding of how social media could be advantageous. They were trying to get to know what social media is and why should the business firms care about it. Unlike traditional media, that provides a one-way experience (in which media broadcast some news about the business for public consumption), social media provides a two-way experience of interaction where consumers can interact with companies. Customers in touch with social media, unlike customers of traditional media, can interact immediately and directly with either the author or the originator of the proffered information. The cross communication and interaction that social media makes possible is the reason what makes social media as a tool for changing the strategies in the business that aids towards achieving business targets.

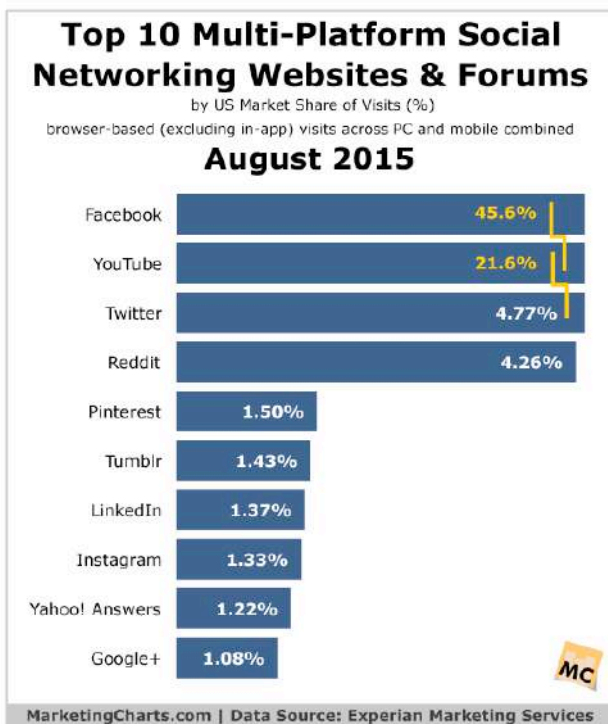


Figure 1 Famous Social Media Sites and Usage of them

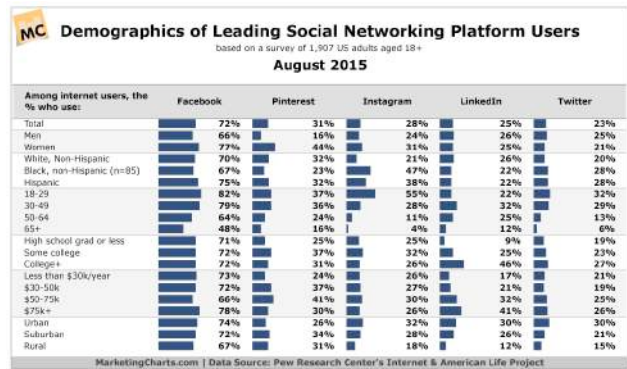


Figure 2 Leading Social Networking Platform Users

Figure 1 and 2, elaborates the demand and usage of social media among the people. That would be an ideal example to show the extent to which people could be reached through social media.

C. How Social Media aids to Competitive Advantage

Michael Porter's Value Chain Analysis

One way to explore about the impact of social media within a business firm is to know whether the organization will be able to attain competitive advantage with its customers.

In this review article, we will use the concept of competitive advantage as proposed by Porter in his work in 1985. Michael Porter has described how companies can develop sustainable competitive advantage in his publication in 1985. Porter describes three different ways that firms can use for competitive advantage. They are Cost leadership, Differentiation and Focus. Moreover, according to Porter's model, a firm's Value Chain composed of all the activities that are required for the production of goods and services. These activities can be major milestone activities that add value directly to the produced goods or services and provide support or secondary activities that add indirect value to end product.

A business firm can find out the ways of gaining competitive advantage by analysing its value chain. This analysis is based on finding the impact of both primary and secondary activities on any things of Cost Leadership, Differentiation Advantage or Focus. Based on this analysis, the activities identified can be improved so that it leads towards achieving competitive advantage or removing competitive disadvantage.

We use the same logic to social media. In order to utilize social media to gain competitive advantage, organization's value chain activities that can be impacted by social media are identified (Atzmueller & Martin, 2011). The weight of impact can be measured. This can be

analysed for focus, cost, as well as differentiation advantage.

List of social media activities that can be performed in each primary activity can be proposed as follows. Similar activities can also be extended to secondary activities as well.

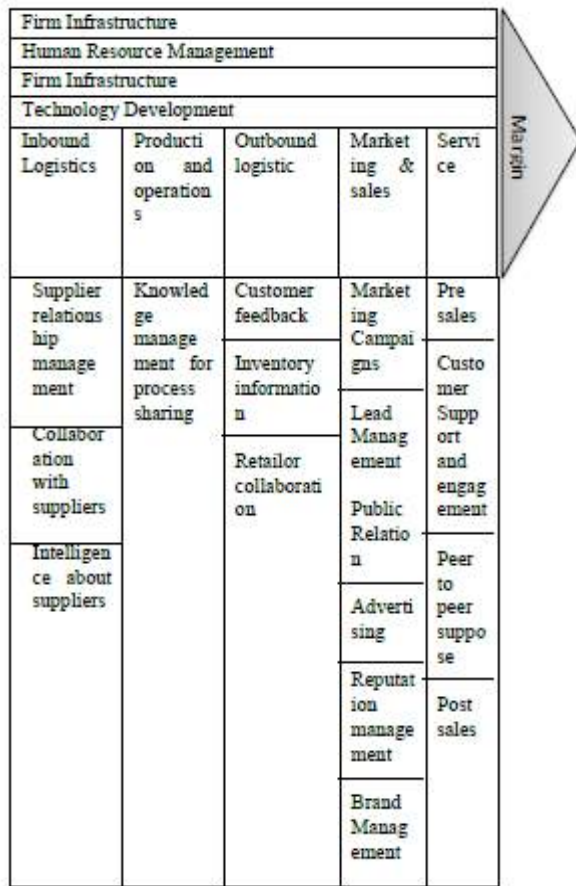


Figure 3 Usage of social media in Porter's generic value chain

Out of all the generic value chain activities, sales and marketing makes maximum out of social media. Fashion retailers like Myntra.com and food sites like Foodpanda.in regularly use social media channels to run marketing campaigns and offer discounts (Singla, M.L & Durga, A , 2015). In fact marketing campaigns does not only focus on offering discounts. Many organizations segment the users and target them on social media with various offerings as well as to collect feedback. Users who show more interest with these brands their offerings become influencers and customers. From that way, social media acts as an intermediate between consumer and the firm to lead acquisition channel to get more leads that are later managed in a Customer Relationship Management (CRM) system.

Value Chain Activity	Social Media Use Case	Social Media Technologies
Inbound Logistics	Supplier relationship management	Social Networks
	Collaboration with suppliers	Wiki, Blog, KM, YouTube
Operations	Intelligence about suppliers	Social Media Analytics, Twitter, LinkedIn, Blogs
	Knowledge Management for process sharing	Wiki, Blog, Forums, Document Sharing
Outbound Logistics	Customer feedback	Twitter, Facebook, Blogs
	Inventory information	Twitter
Sales and Marketing	Retailer collaboration	Wiki, Blog, KM, YouTube
	Marketing Campaigns	Social Media Marketing
	Lead Management	Twitter, Facebook, Blogs
	Public Relation	Twitter, Facebook, Blogs
	Advertising	Twitter, Facebook, Blogs
Sales and Marketing	Reputation Management	Social Media Analytics, Twitter, LinkedIn, Blogs
	Brand Management	Twitter, Facebook, Blogs
Service	Pre-sales	Wiki, Document Sharing, KM
	Customer Support and Engagement	Social Media Marketing, Monitoring and Analytics
	Peer to Peer Support	Wiki, Forums
	Post Sales	Wiki, Forums
Procurement	Hiring (information gathering about potential candidates)	Blog, LinkedIn, Social Media Analytics
	Product comparison including costing	Social Media Analytics, Blogs
Human Resource Management	Organization policy sharing	Knowledge and Document Management
	Recruitment	Blogs, Twitter
Infrastructure	Brand Management	Intra.net
	Employee recognition	
Technology	Document Management for contracts and other documents	Social Content Management, Ratings, Feedback
	Innovation Management	Collaboration
	General collaboration	Collaboration

Figure 4 Social Media Usage and its impact on different sources of competitive advantage

Business Goal	Blog	Podcast	Video	Social Network	Private Community	Recommendation Engine
Build Customer Community	*	*	*	*	*	*
Counter Negative Publicity	*	*	*	*	*	*
CRM's Management	*	*	*	*	*	*
Customers Conversation	*	*	*	*	*	*
Expose Employee Talent	*	*	*	*	*	*
Generate Website Traffic	*	*	*	*	*	*
Humanize The Company	*	*	*	*	*	*

Figure 5 Social Media Tools Used to achieve business goals

D. Key features to be extracted from big data

1) Keyword searching

Organizations can define the topics to be monitored such that it suits their brand images, products and management features. Also keywords targeting the competitive companies also can be defined so that the company achieve competitive strategies over them.

2) Thematic and sentiment analysis

Data available in social media are not organized in a proper way and it's increasingly becoming tedious for businesses to capture, analyse, and make use of it in a timely manner. However, the unstructured data is not useful at all.

As a result, there is a need to study how organisations can (a) extract and analyse social media data related to their business (Sensing), and (b) utilise external intelligence gained from social media monitoring for specific business initiatives (Seizing) (Zhang et al., 2014).

3) *Analysis of spread patterns*

Monitoring is also used to know of how messages spread in social media. One main approach is to find out causal relations and follow the impact of the interventions of social media. Another one is to follow the spread pattern that makes it able to predict its development. Conversations in social networking sites with the most potential for generating and influencing purchasing decisions were monitored using the analytical tools available on the sites to analyse the spread and impact of social media messages (Kumar & Mirchandani, 2012).

4) *Share of Voice*

This is referred to as measuring the work that is surrounded by your business firm and your competitors online. By evaluating the presence of other firms within relevant market place, companies can well understand the success of their own business and marketing activities.

If the competitor has a greater influence of the voice, it may be time to increase your online presence.

5) *Sentiment*

By evaluating the sentiments in comments of data available in social media, businesses can understand the negative points consumers have in mind about their product or service. Say if a competition company is not acknowledging for customer complaints, then you can go into that customer and try to introduce our products.

Also by getting to know needs and wishes of a consumer within the industry, a business can try to avoid some mistakes and try to improve them in their own business strategies (Eman & Younis, 2015).

6) *Top Sites and Page Types*

It is very much important to know where our competitors are active online. Firstly, it may let you know some important sites or forums where a competition brand is dominating its presence, that your firm not aware of. This information will be useful to understand whether you also need to get into that platform to increase your strength or to be a leader in the industry.

Another advantage of finding out that is to get to know where the competitor companies are not actively participating. It would be useful to know if that platform is not at all needed or if that platform is the best place to place your footprint so that without competition you can be a leader there.

7) *Influencers*

By monitoring social media, you will be able to find out influencers, who normally criticize about your product or

services or else your competitor's. By identifying that you will be able to correct your weaknesses so that consumers will be satisfied. Or else, by knowing competitor's weakness you will be able to give a better product or service to your customers over the competitor.

8) *Topics*

This is finding out the online conversations about our business or the competitor.

By finding out this you get to know about the emerging trends in the business firms which your competitor is using or you may use that to be the first person to provide services using emerging trends. By finding out the new trends, you can become the market lead.

9) *Audience Type*

By analysing about the people who are discussing about your product, you will get a clear understanding about the interested people and therefore get an idea about audience whom you want to target. Say for an example, you get identify which age group is mostly interested on your products, so that you can develop strategies according to it.

E. Development methods in social media monitoring

In academia, various methods to navigate, collect, analyse the content, sensing and social media topics have been proposed. Many techniques categorization unstructured data were adapted and implemented in the social media studies and areas of knowledge. These efforts require interdisciplinary research skills, since they involve the treatment of raw data regarding the search for information quality and significance as displayed, most appropriate to the situation. Reweaves points out that due to the volume and complexity of data, simple models are inadequate and, therefore, the process of data mining of social media requires a comprehensive approach, highlighting the need for a unified approach to explore the data structure for efficient identification, classification and forecasting topics, issues and relevant models. Countless techniques such as extraction, sorting or categorización, feature analysis, linguistic analysis, content analysis, association between texts, grouping and summary can be used to extracted knowledge models or interest sentences and unexpected textual documents. Researchers say the clustering technique offers the advantage of revealing trends, unexpected correlations or patterns in the structure of the data that had not been budgeted. Scholars emphasize that the cluster analysis is a technique well explored in data mining and, according to Abdous, He & Yen (2012) explored technique that allows display models, grouping similar words and terms or values which are also attributed & similarly coded. The same authors say that even a perspective of data mining,

clustering also can be used for this. Introduction of Artificial intelligence techniques to monitor social media is also a progress in the development of social media monitoring. Some of those type techniques are:

1) Text Mining

Data mining, text mining activities as a source of data that accepts text work is intended to gather data on other defined structured text. For example, the classification of text, segmentation of, subject extraction from text, production of class particle, emotional analysis, text summarization, and entity relationship modeling aims such studies.

Bringing information within the scope of the text mining efforts in order to achieve the above goals (information retrieval), syllable analysis (lexical analysis), word frequency distribution, pattern recognition, information Extraction, data mining and even visualization (Wei Jin *etal*).

Text mining studies aim to achieve results through the text more statistically. Often for feature extraction using natural language processing, text mining studies are used.

Introduced in 1995 by Ronen Feldman and Ido Dagan as "Knowledge Discovery from Text (KDT)". In the research terminology text mining is not a clearly defined term. In analogy to data mining in the Knowledge Discovery in Databases (KDD) Text Mining is a largely automated process of knowledge discovery in textual data, which is to enable an effective and efficient use of available text archives. Comprehensively we can say text mining as a process of compilation and organization of the formal structuring and algorithmic analysis of large collections of documents for the appropriate extraction of information and the detection of concealed content relationships between texts and text fragments.

The different views of text mining can be sorted by various typologies. These kinds of information retrieval (IR), the document clustering, the text data mining and KDD are repeatedly referred to as sub-forms of text mining.

Most similar to the text mining, data mining with this, it shares many methods, but not the subject matter: While data mining is usually applied to highly structured data, text deal mining with significantly weaker structured text data. In text mining, the primary data are therefore in a first step, more structured to allow its development with methods of data mining. Unlike most tasks of data mining, multiple classifications are usually welcomed in the text mining (Miloš & Milana, 2008).

2) Web Mining

Web mining is a process used for meaningful patterns from Web resources, profiles, physician data mining techniques (to find a useful correlation is hidden among other data, to find such a trend Extract the actionable information for future decisions) to be applied once. Its application fields include information filtering and monitoring of such competitors as patents and technology development, mining utilization of the web access logs for analysis, browsing (breadcrumb navigation in your web), etc. support.

Web mining is part of the data mining specialist for disclosure of information on the Internet, especially the World Wide Web. Huge databases are rich in data, as well as poor information that is hidden in stored data. Data mining helps to reveal important information and knowledge applied to the data.

Also, greatly contributes to decision-making, business and science. Mining the Web is used by a variety of companies (personalized marketing), as well as governmental organizations (fight against terrorism). Companies can establish better relationships with their customers because they can better meet the needs of consumers, and react more quickly to their queries. There are also problems in mining web. The technology Web mining does not cause any problems, while the use of these technologies in the personal data can do damage. The largest such issue would be invasion of privacy. Privacy is violated when information about an individual obtained, used and disseminated without his knowledge or permission. Another such problem is that companies that collect data for a specific purpose, such data can be used for something completely different. The truth is that these companies are responsible for the issuance of all of this data, and if you notice some irregularities followed by a lawsuit, but there is no law that prevents them from trading in such data (Sheibani & Amir). As the popularity of the World Wide Web continues to grow, there is a growing need to develop new tools and techniques that will improve the overall usefulness.

3) Named Entity Recognition

The named entity recognition is a subtask of information extraction activity in the documentary corpus. It is to look for text objects (that is to say a word or group of words) be categorized into classes such as personal names, names of organizations or businesses, place names, quantities, distances , values, dates, etc.

For example, you could give the following text, labeled by an entity named recognition system used in a survey.
Jim went to see Kim at 9:00 of May 18.

This text can be recognized as
 <PERSON>Jim</PERSON> went to see
 <PERSON>Kim</PERSON> at <TIME>9:00</TIME> of
 <DATE>May 18</DATE>.

The text of this example is tagged with XML tags. Most labeling systems use formal grammars associated with statistical models, possibly supplemented by databases (lists names, names of cities or countries, for example). In large surveys, manually written grammars database systems perform best. The downside is that such systems sometimes require months of drafting work.

Current statistical systems use to share a large amount of pre-annotated data to learn the possible forms named entities. There is no need here to write many rules in hand, but to label a corpus that will serve as a learning tool. These systems are also very costly in human time. To solve this problem, recently, initiatives such as DBpedia Yago or seek to provide semantic corpus that could help design labeling tools. In the same spirit, some semantic ontology such as NLGbAse is largely oriented labeling. Since 1998, the annotation of named entities in texts met with increasing interest. There are many applications call for finding information or, more generally understanding of textual records.

There is not strictly speaking labeling standards. Labels are heavily skewed depending on the application needs: we usually find the label type classes roots Person, Organization, Product, Place, plus the label length and quantity (time and amount).

F. How these techniques are implemented in social media monitoring

Social Media monitoring can be broadly defined as using Social Media data sources, many of which can be taken automatically using some extracting tools, and choosing and aiming concepts and word types to analyse market behaviour and interaction(Sponder & Marshall, 2012). Thus, data collected from this sort can be summarized, analysed and classified and passed through various types of algorithms to get patterns and trends that are useful to achieve business goals. In order to do this, the social media need to be monitored on a specific time basis and filter the useful things and wan to decide which is worth to consider upon. This is known as social media monitoring.

Monitoring social media using text mining has been proposed by various researchers as depicted in following pictures.

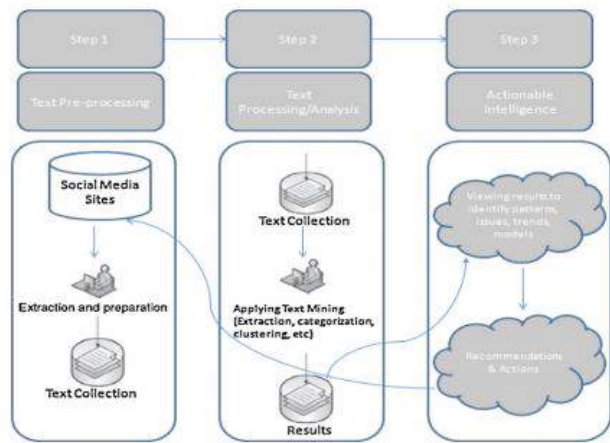


Figure 6 Implementing text mining

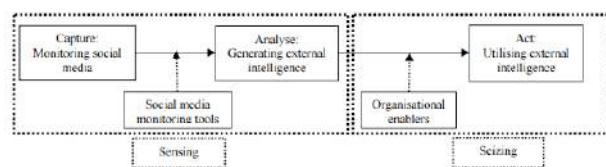


Figure 7 Big data analysis of Social Media

III. CONCLUSION AND RECOMMENDATIONS

Most business firms, government agencies, non-profit organizations (e.g.: Hospitals, universities, associations, etc.) and individuals, such as writers and celebrities use media monitoring as a tool for identifying mentions of their organization, their brands and leaders in news media. some organizations also provide media monitoring tools to track the success of their messages to find information about competitors and relevant specific issues on the organization to manage the benchmark performance against competitors, corporate or brand image, to gather the information industry to create a better to understand the strengths and weaknesses of corporate communications, to identifying new business opportunities, and for other purposes.

Despite the initially purely private use of social networks were in the past few years an increasing number of private companies about to adapt their presence on the Internet to Web 2.0 developments to benefit from the advantages. Primarily social media provide platforms for companies, a relatively inexpensive but accurate marketing tool because of the wastage of classical forms of marketing can be significantly reduced by targeting of customers and potential customers. Furthermore, the operation of a social media concept supports the development of new target groups, mainly the move on the internet and can hardly be achieved through traditional channels. In this context, especially the resulting economic potential of viral marketing is to be mentioned the results from the typical social media concept of networking between users.

Rather than only monitoring social network sites, I recommend that the use of internal networks can also be good for performance improvement of a company which leads to gaining competitive advantage. Such a system is not only used to improve the working climate by facilitating and more informal communication among the employees, but also helps to uncover organizational weaknesses and eliminate them.

ACKNOWLEDGMENT

Our special thank goes to the academic staff of the Department of ICT of the South Eastern University of Sri Lanka for their continuous support given for us for working on this research. Also a heartfelt thank is dedicated to the team "Business Intelligence" of IFS R&D International Pvt Limited for supporting to make this research work a success.

REFERENCES

Wei Jin, Hung Hay Ho and Rohini K. Srihari. 2009. Opinion Miner: a novel machine learning system for web opinion mining and extraction. Proc. of the 15th ACM SIGKDD, 1195–1204.

Eman M.G. Younis . 2015.Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study, Faculty of Computer and Information Minia University, Egypt International Journal of Computer Applications (0975 – 8887) Volume 112.

Riswana Irfan. 2004. A Survey on Text Mining in Social Networks, The Knowledge Engineering Review, Vol. 00:0, 1–24. Cambridge University Press DOI: 10.1017/S0000000000000000 Printed in the United Kingdom.

Miloš Radovanović and Mirjana Ivanović. 2008. TEXT MINING: APPROACHES AND APPLICATIONS, Vol. 38, No. 3 , 227-234.

Abdous, M and Yen,H.. 2012.Using Data Mining for Predicting Relationships between Online Question Theme and Final Grade. Educational Technology & Society, 15 (3), 77–88.

Ruggiero, A and Vos, M. 2014. Social Media Monitoring for Crisis Communication: Process, Methods and Trends in the Scientific Literature. Online Journal of Communication and Media Technologies Volume: 4 – Issue: 1.

Atzmueller, Martin. Modeling and Mining Ubiquitous Social Media International Workshops MSM 2011, Boston, MA, USA, October 9, 2011, and MUSE 2011, Athens, Greece, September 5, 2011, Revised Selected Papers. Berlin: Springer, 2012.

Fang, Xing, and Justin Zhan. "Sentiment Analysis Using Product Review Data." Journal of Big Data (2015). Han, Jungkyu, and Min Luo. "Bootstrapping K-means for Big Data Analysis." 2014 IEEE International Conference on Big Data (Big Data).

He, Wu, Shenghua Zha, and Ling Li. "Social Media Competitive

Analysis and Text Mining: A Case Study in the Pizza Industry." International Journal of Information Management: 464-72.

Hu, Xia, and Huan Liu. "Text Analytics in Social Media." Mining Text Data: 385-414.

Meyrowitz, Joshua. No Sense of Place: The Impact of Electronic Media on Social Behavior. New York: Oxford UP, 1985.

"Patent Issued for Signal Processing Approach to Sentiment Analysis for Entities in Documents." Journal of Engineering 28 May 2014.

Rastunkov, Vladimir, and Mark Rusch. "Using STATISTICA Text Miner to Monitor and Predict Success of Marketing Campaigns Based on Social Media Data." Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications: 151-80.

Sheibani, Amir A. "Opinion Mining and Opinion Spam: A Literature Review Focusing on Product Reviews." 6th International Symposium on Telecommunications (IST).

Sponder, Marshall. Social Media Analytics: Effective Tools for Building, Interpreting, and Using Metrics. New York: McGraw-Hill, 2012.