



# Mining for Social Media: Usage Patterns of Small Businesses

*Shilpa Balan, Janhavi Rege*

*California State University-Los Angeles, College of Business and Economics*

## Abstract

**Background:** Information can now be rapidly exchanged due to social media. Due to its openness, Twitter has generated massive amounts of data. In this paper, we apply data mining and analytics to extract the usage patterns of social media by small businesses. **Objectives:** The aim of this paper is to describe with an example how data mining can be applied to social media. This paper further examines the impact of social media on small businesses. The Twitter posts related to small businesses are analyzed in detail. **Methods/Approach:** The patterns of social media usage by small businesses are observed using IBM Watson Analytics. In this paper, we particularly analyze tweets on Twitter for the hashtag #smallbusiness. **Results:** It is found that the number of females posting topics related to small business on Twitter is greater than the number of males. It is also found that the number of negative posts in Twitter is relatively low. **Conclusions:** Small firms are beginning to understand the importance of social media to realize their business goals. For future research, further analysis can be performed on the date and time the tweets were posted.

**Keywords:** social media, analysis, data mining, small business

**JEL classification:** C2, F23, L82

**Paper type:** Research article

**Received:** Jan 15, 2017

**Accepted:** Mar 31, 2017

**Citation:** Balan, S., Rege, J. (2017), "Mining for Social Media: Usage Patterns of Small Businesses", *Business Systems Research*, Vol. 8, No. 1, pp. 43-50.

**DOI:** 10.1515/bsrj-2017-0004

## Introduction

Social media are computer-facilitated tools that enable the faster exchange of information in virtual networks (Buettner, 2016). Millions of users use social media websites every day. The most widely used social media websites are Facebook, WhatsApp, Instagram, Twitter, and YouTube. Social networking websites are now primary means of communication for people among all age groups. In this paper, we apply data mining and analytics to extract the usage patterns of social media by small businesses. Specifically, we look at the Twitter posts related to small business. Businesses in the United States that have lower than 500 employees are defined as small business and they represent 99.7% of all firms. Hence, it is vital to study how small businesses are using social media as a means to promote their business.

In a business setting, social media analytics is a subset of Business Intelligence (BI) that is focused on methodologies and technologies that transforms unstructured

data from social media into meaningful information for business purposes (Stieglitz et al., 2014). Business Intelligence (BI) refers to techniques used for analyzing data (Cebotarean et al., 2017).

Data analytics are critical in supporting decision making. Analytics is helping organizations to reduce costs. Social media offers opportunities for social communications for a business (Fischer et al., 2011). Social media analytics can be applied to understand the user sentiments about a company or a product (Mosley, 2012). To make this sort of prediction, business analytics and intelligence is applied to this research. Huff (2015) mentions that business intelligence can allow companies to save time and to focus resources on more profitable opportunities. BI tools can predict future outcomes based on historical data.

According to Chesbrough (2003), small businesses can use social media to reach out to customers thus increasing their revenue. Apenteng and Doe (2014) state that a survey from LinkedIn displayed that 80% of small businesses are moving to social media sites.

Cardon and Marshall (2015) state that social networking sites are now more commonly used for online communication than email. As social media has started to impact people's lives, a study of its usage would thus be significant. In this research, the patterns of social media usage for Twitter posts on small business are examined using IBM Watson Analytics. Further, sentiment analysis is applied to the data. Sentiment analysis refers to "the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials" (Tejwani, 2014, p1). The following research questions are examined in this study:

1. Is there a gender difference in social media usage for small businesses?
2. What is the sentiment analysis of social media users for small businesses? Is the sentiment of the small businesses users using social media positive or negative or neutral?
3. Is there a relationship between the small business posts on social media and the language in which the posts are made?
4. Which state in the United States uses social media the most for small businesses?

## Research Background

Social media has impacted customer perceptions and decisions. Kaplan and Haenlein (2010) state that social media has generated massive content due to users sharing their opinions and experiences on several topics.

Small enterprises can use social media such as forums and blogs to build relationships (Eid et al., 2013). They can also use crowdsourcing opportunities (Howe, 2006) for technical problems.

Analytics can be applied to social media to identify expressions that provide insights to the social media posts. Currently, research efforts have been focused on issues toward understanding social media. For example, in social media data analysis, companies see the opportunity in advertising, and social customer relationship management (Coen, 2016).

Social media is a means to connect to a larger audience. It promotes social capital (Hetz et al., 2015), which is the network of relationships among people. According to Kiron et al. (2013), social media is a low-cost means to increase customer awareness of products and services.

Currently, many firms have huge quantities of data generated due to the presence of social media. Social media is one of the major data sources for many

companies. Companies need to analyze these huge quantities of data to manage customer preferences to improve the firm's performance and to stay ahead of competition.

Small businesses are able to reach a larger market on the Internet (Community Futures PA and Districts, 2012, Merrill et al., 2011). With the help of social media, businesses are able to increase the growth in their number of customers. For example, when one customer or client selects the 'like' button of the firm's page on Facebook, it is shown on the customer's friend's feeds. This could lead to potential customers.

The use of social media provides the insight about a buyer's dynamics. Some scientists from the University of Milan and the Massachusetts Institute of Technology (MIT) found that users display physical and psychophysiological reactions when they log onto Facebook.

One popularly used social media website is Twitter. Twitter is a social networking website that allows users to send short messages. According to Felt (2016), Twitter is one of the largest social media platforms. Twitter was created in 2006 (Mosley, 2012). The growth of Twitter has been extraordinary. As of 2011, there were 200 million registered accounts in Twitter (Banking.com, 2011). Twitter is appealing due to the openness of the data. Burgess et al. (2015) noted that the openness of Twitter has created an enormous amount of social media data.

In this paper, usage patterns of Twitter posts on small businesses are examined using IBM Watson Analytics. Watson Analytics is a cloud-based service intended to provide the ability to analyse data without much difficulty. Further, a sentiment analysis is conducted on the data.

## Methodology

As stated previously, in this research, IBM Watson Analytics is used to observe the patterns of social media usage on Twitter posts related to small business. In this paper, we particularly analyse tweets on Twitter for the hashtag #smallbusiness. The tweets from Twitter were extracted using IBM Watson Analytics. The dataset extracted from Twitter is from January 2017 to February 2017.

The data extracted from Twitter contains information about authors, their posts, their genders, the location from where the tweets are posted, and the time at which they have been posted.

The data analysis performed in this research using IBM Watson Analytics provides insights on the usage patterns of Twitter by small business users. Further, the sentiment analysis of the data is also analysed using IBM Watson Analytics. The analysis results are described in the Results and Analysis section. Before analysing the data, the unstructured data extracted from Twitter had to be cleaned and organized.

### *Data Cleaning and Data Refinement*

This section describes the data cleaning and refinement performed on the data sets used for this research. Most often, data involves unnecessary or ambiguous content. Any ambiguous fields and empty or missing values in the data need to be managed appropriately. For example, ambiguous fields could be deleted from the data set if it is not possible to edit these fields with accurate data.

The data set needs to be refined to make it more meaningful for analysis. Before analysing the data, we examined the data to ensure it was cleaned. Following are the primary steps we performed for data cleaning:

- Null values: We examined the Twitter data set extracted for any null or ambiguous values. We did not find null or ambiguous values in the data set.
- Missing values: We examined the Twitter data set to check for missing values. For the twitter data analysed, we did not find any missing values.
- Identifying relevant columns for analysis: Any columns not used for analysis were removed from the dataset. Only the columns required for analysis are retained. For example, the time the tweet was posted was not taken into consideration. Also, the author location was not taken into consideration. For the analysis in this study, we only analysed the tweets and its sentiment.

After refining the data, the quality of the data for analysis improved.

### Correlation Analysis

Mosley (2012) applied Cramer's V statistic to find the association between two variables. In this paper, we apply the Cramer's V statistic for pairs of hashtag associations using Mosley's methodology. To calculate the Cramer's V statistic, we use similar methodology as that of Mosley (2012). Consider a 2 x 2 matrix indicating the frequency of the combination of two hashtags. Cramer's V formula is as follows:

$$Cramer's V = \frac{n_{00} n_{11} - n_{01} n_{10}}{\sqrt{n_{00} n_{11} n_{01} n_{10}}}$$

The representation  $n_{ab}$  symbolises the frequency of the combination of hashtags in the dataset. For example,  $n_{00}$  is the number of tweets where neither hashtag1 nor hashtag2 were present.  $n_{.b}$  is the frequency for column b, while  $n_{.a}$  is the frequency for row a.

The result ranges between -1 and 1, where a value of -1 indicates a negative correlation, a value of 1 indicates a perfect positive correlation, and 0 indicates no correlation.

## Results and Analysis

This section describes the findings of our analysis for the Twitter hashtag #smallbusiness. Table 1 lists some hashtags that were co-associated to the hashtag #smallbusiness. For example, hashtags #website, #entrepreneur, #webhosting and #blog were most co-associated to the hashtag #smallbusiness.

Table 1

Cramer's V statistic for sample co-associated hashtags with the hashtag #smallbusiness

Co-associated Hashtags with #smallbusiness	Cramer's V statistic
#website	0.85
#entrepreneur	0.65
#webdesign	0.54
#webhosting	0.61
#startup	0.59
#mobile	0.44
#google	0.34
#blog	0.67

Source: Author's work

Table 2 shows that most of the tweets for small businesses are posted by females. 72.36% of the tweets on small business are posted by females while only 27.63% of the tweets are posted by males. This shows that women participate in small business discussions more than men. The most of the tweets related to small business were posted in English and Spanish. Also, the number of neutral tweets is more than the number of positive or negative tweets for the hashtag #smallbusiness.

Table 2  
Tweet Count for the Hashtag #smallbusiness

	Characteristics	Author Tweet Count percentage for #smallbusiness
<b>Gender</b>	Male	27.63%
	Female	72.36%
<b>Language</b>	Spanish	47.53%
	English	40.06%
	Dutch	6.73%
	French	5.67%
<b>Author Sentiment</b>	Positive	22.72%
	Negative	4.55%
	Neutral	72.72%

Source: Author's work

Table 3 shows that the most number of tweets related to small business in the United States were posted in the state of South Dakota.

Table 3  
Top 9 states in United States by Tweet count for the Hashtag #smallbusiness

Author State	Ranking of State in United States by Author Tweet Count for #smallbusiness
South Dakota	1
South Carolina	2
Minnesota	3
New Jersey	4
California	5
New York	6
Pennsylvania	7
Massachusetts	8
North Carolina	9

Source: Author's work

## Discussion

Businesses are using social media as well. Every post or tweet on a social media site may be valuable to companies. Del Rowe (2016) states that analytics is a core component of all major decisions in business as it could help in achieving specific business goals. Table 4 shows a sample list of positive and negative words extracted and compiled for the hashtag #smallbusiness. However, the overall number of negative posts was low as seen in Table 2. Small businesses can gain enormous value from social media data. For example, for businesses, social media analytics can help to understand customer preferences on various products.

Table 4

Example of positive and negative words compiled for the hashtag #smallbusiness

Serial No.	Positive words	Negative words
1	Amazing	Avoid
2	Authentic	Mistakes
3	Best	Bad
4	Benefits	Complicated
5	Better	Error
6	Great	Fail
7	Happy	sad
8	Inspiring	Unhappy
10	Productive	
11	Thankful	

Source: Author's work

## Conclusion and Future Research

Small businesses are increasingly using social media. This growth in social media has increased the amount of data generated and this can provide further insights to companies.

Fluss (2013) states that social media is going to change the business setting for many organizations within the next few years. This is because the volume of comments and posts on social media sites is expected to grow rapidly. Further, Fluss (2013) mentioned that those organizations that will invest in incorporating social media will have a major advantage over their competitors.

Social media is important for all businesses because it allows businesses to easily communicate with their customers (Grewal et al., 2013; Smith et al., 2011). However, social media has been extremely more important for small businesses because these businesses lack the resources to market their products or services (Barnes et al., 2012; Levy et al., 2003). The internet has increased the ability of the organizations as well as the customers to connect with one another (Taneja et al., 2014).

In the current social media driven setting, it is vital that small businesses understand the strategies in using social media. Social media marketing helps firms to better understand their customer needs. To maximize the number of people a business can reach, a business must have a social media presence. It is found that Twitter is a good medium for reaching out to people and around 29% of small businesses stated that social media marketing is the primary focus area.

Social media provides businesses the opportunity to communicate with their customers. Generally, customers do not want to follow a business that only posts about their products without interactivity with the customers. Businesses should include photos and videos of their product instead of simply posting messages about their business. One of the ways to engage customers is by asking the customers for their feedback on the products.

Small firms are beginning to realize the importance of social media for their business purposes. For future research, further analysis can be performed on the date and time the tweets were posted. This may provide further analysis of the social media usage patterns for small businesses. In addition, the number of retweets can be taken into consideration while performing sentiment analysis.

## References

1. Apenteng, S. A., Doe, N. P. (2014), "Social Media & Business Growth: Why Small/Medium-Scale Enterprises in the Developing World Should take an Advantage of it", *Journal of Business and Management*, Vol. 16 No. 5, pp. 76-80.
2. Barnes, D., Clear, F., Dyerson, R., Harindranath, G., Harris, L., Rae, A. (2012), "Web 2.0 and micro-businesses: an exploratory investigation", *Journal of Small Business and Enterprise Development*, Vol. 19 No. 4, pp. 687-711.
3. Banking.com (2011), "Social Media Statistics: By-the-Numbers", available at <http://banking.com/analysis/social-media-statistics-by-the-numbers-february-2011/> (14 November 2016)
4. Buettner, R. (2016), "Getting a Job via Career-oriented Social Networking Sites: The Weakness of Ties", 49th Hawaii International Conference on System Sciences (HICSS-49), January 5-8, 2016, Kauai, Hawaii.
5. Burgess, J., Bruns, A. (2015), "Easy data, hard data: The politics and pragmatics of Twitter research after the computational turn", in Langlois, G., Redden, J., Elmer, G. (Eds.), *Compromised Data: From Social Media to Big Data*, New York, Bloomsbury, pp. 93-111.
6. Cardon, P. W., Marshall, B. (2015), "The hype and reality of social media use for work collaboration and team communication", *International Journal of Business Communication*, Vol. 52 No. 3, pp. 273-293.
7. Cebotarean, E., Titu, M. (2017), "Business Intelligence", *Journal of Knowledge Management, Economics and Information Technology*, pp. 1-12.
8. Chesbrough, H. (2003). *Open Innovation: The New Imperative for Creating and Profiting from Technology*, Boston, Harvard Business School Press.
9. Coen, B. (2016), "Using Social Media for Customer Relationship Management", available at: <http://www.socialmediatoday.com/social-business/using-social-media-customer-relationship-management/> (20 December 2016)
10. Community Futures PA and Districts (2012). *An Introduction to Social Media for Small Businesses*. Blue Beetle Books.
11. Del Rowe, S. (2016), "Take an Adaptive Approach to Analytics: Analytics focused on Business Objectives Should be the Basis of Business Decisions", *CRM Magazine*, Vol. 20 No. 5, p. 19.
12. Eid, R. E., El-Gohary, H. (2013), "The impact of E-marketing use on small business enterprises' marketing success". *The Service Industries Journal*, Vol. 33 No.1, pp. 31-50.
13. Felt, M. (2016), "Social media and the social sciences: How researchers employ big data analytics", *Big Data & Society*, Vol. 3 No.1, pp. 1-15.
14. Fischer, E., Reuber, A. R. (2011), "Social interaction via new social media: (How) can interactions on Twitter affect effectual thinking and behavior?", *Journal of Business Venturing*, Vol. 26 No. 1, pp. 1-18.
15. Fluss, D. (2013), "Meeting the Social Media Challenge", *CRM Magazine*, Vol. 17 No. 9, pp. 38-39.
16. Grewal, D., Levy, M. (2013). *Marketing*, New York, McGraw-Hill Publishing.
17. Hetz, P. R., Dawson, C. L., Cullen, T. A. (2015), "Social Media Use and the Fear of Missing out While Studying Abroad", *Journal of Research on Technology in Education*, Vol. 47 No. 4, pp. 259-272.
18. Howe, J. (2006), "The Rise of Crowdsourcing", available at: <https://www.wired.com/2006/06/crowds/> (27 February 2017)
19. Huff, A. (2015), "TMW Systems shows fleets how to make time with business intelligence", *Commercial Carrier Journal*, Vol. 172 No. 10, pp. 41-42.

20. Kaplan, A. M., Haenlein, M. (2010), „Users of the World, Unite! The Challenges and Opportunities of Social Media”, *Business Horizons*, Vol. 53 No. 1, pp. 59-68.
21. Kiron, D., Ferguson, R. B., Prentice, P. K. (2013), “From Value to Vision: Reimagining the Possible with Data Analytics”, *MIT Sloan Management Review: Spring Research Report*, pp. 1-19.
22. Levy, M., Powell, P. (2003), “Exploring SME Internet adoption: towards a transporter model”, *Electronic Markets*, Vol. 13 No. 2, pp. 173-181.
23. Merrill, T., Latham, K., Santalesa, R., Navetta, D. (2011), “Social Media: The Business Benefits May be Enormous, But Can the Risks - Reputational, Legal, Operational - Be Mitigated?”, available at: [http://old.findinet.nl/~uploads/newsModule/ace\\_socialmedia111118.pdf](http://old.findinet.nl/~uploads/newsModule/ace_socialmedia111118.pdf) (21 April 2017)
24. Mosley, R. (2012), “Social Media Analytics: Data Mining Applied to Insurance Twitter Posts”, *Casualty Actuarial Society E-Forum*, Vol. 2, pp. 1-36.
25. Smith, N., Wollan, R. (2011). *The power and business risks of social media. The Social Media Management Handbook: Everything You Need to Know to Get Social Media Working in Your Business*, Hoboken, John Wiley & Sons.
26. Stieglitz, S., Dang-Xuan, L., Bruns, A., Neuberger, C. (2014), „Social media analytics: An interdisciplinary approach and its implications for information systems”, *Business & Information Systems Engineering*, Vol. 6 No. 2, pp. 89-96.
27. Taneja, S., Toombs, L. (2014), “Putting a face on small businesses: Visibility, viability, and sustainability the impact of social media on small business marketing”, *Academy of Marketing Studies Journal*, Vol. 18 No. 1, pp. 249-260.
28. Tejwani, R. (2014), “Sentiment Analysis: A Survey”, *ArXiv*, pp. 1-3.

## About the authors

Shilpa Balan is an Assistant Professor in the Department of Information Systems, College of Business and Economics at California State University-Los Angeles. Her main research interests include business intelligence, social media mining, social network analysis and big data analytics. The author can be contacted at [sbalan@calstatela.edu](mailto:sbalan@calstatela.edu)

Janhavi Rege is a MBA student with a focus on Business Analytics in the College of Business and Economics at California State University-Los Angeles. Her main research interests include business intelligence and social media mining. The author can be contacted at [jrege@calstatela.edu](mailto:jrege@calstatela.edu)