TEXT MINING AND SOCIAL MEDIA ANALYSIS OF PIZZA INDUSTRY USING R

Mr. Shaikh Mudassir Nadeem[1]

[1] *Mr. Shaikh Mudassir Nadeem, Department of Computer Science & Engineering, EESCOE&T, Ohar, Aurangabad, Maharashtra, India*

ABSTRACT

Social media have been adopted by many businesses. Many companies are using social media tools such as Facebook and Twitter to provide various services and interact with customers. As a result, a large amount of user-generated content is freely available on social media sites. To increase competitive advantage and effectively assess the competitive environment of businesses, companies need to monitor and analyze not only the customer-generated content on their own social media sites, but also the textual information on their competitors' social media sites. In an effort to help companies understand how to perform a social media competitive analysis and transform social media data into knowledge for decision makers and e-marketers, this paper describes an in-depth case study which applies text mining to analyse unstructured text content on Facebook and Twitter sites of the three largest pizza chains: Pizza Hut, Domino's Pizza and Papa John's Pizza. The results reveal the value of social media competitive analysis and the power of text mining as an effective technique to extract business value from the vast amount of available social media data. Recommendations are also provided to help companies develop their social media competitive analysis strategy.

Keyword: *Social media, Facebook, Twitter, R Studio, Text mining, Social media mining, Competitive analysis, Actionable intelligence.*

1. INTRODUCTION

Data mining is the process of extracting the patterns from large amount of data. Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining. Thus, data mining should have been more appropriately named "knowledge mining from data," which is unfortunately somewhat long. "Knowledge mining," a shorter term may not reflect the emphasis on mining from large amounts of data [1]. Social media is computer-mediated tool that allow people to create, share or exchange information, career interests, ideas, and pictures/videos in virtual communities and networks. Social media is defined as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content. Furthermore, social media depends on mobile and web-based technologies to create highly interactive platforms through which individuals and communities share, co-create, discuss, and modify user-generated content [2].

Social media mining is a process to extract particular patterns from social media sites with the aim of analyzing the data. Following are the different types of social media:

Social Networks – Services that allow you to connect with other people of similar interests and background. Usually they consist of a profile, various ways to interact with other users, ability to setup groups, etc. The most popular are Facebook and LinkedIn.

Bookmarking Sites – Services that allow you to save organize and manage links to various websites and resources around the internet. Most allow you to "tag" your links to make them easy to search and share. The most popular are Delicious and StumbleUpon.
Social News – Services that allow people to post various news items or links to outside articles and then allows its users to "vote" on the items. The voting is the core social aspect as the items that get the most votes are displayed the most prominently. The community decides which news items get seen by more people. The most popular are Digg and Reddit.

Media Sharing – Services that allow you to upload and share various media such as pictures and video. Most services have additional social features such as profiles, commenting, etc. The most popular are YouTube and Flicker.

Microblogging – Services that focus on short updates that are pushed out to anyone subscribed to receive the updates. The most popular is Twitter.

Blog Comments and Forums – Online forums allow members to hold conversations by posting messages. Blog comments are similar except they are attached to blogs and usually the discussion centers around the topic of the blog post. There are MANY popular blogs and forums.

1.1 Social Media Mining

Social Media Mining is the process of representing, analyzing, and extracting actionable patterns from social media data. Social Media Mining, introduces basic concepts and principal algorithms suitable for investigating massive social media data; it discusses theories and methodologies from different disciplines such as computer science, data mining, machine learning, social network analysis, network science, sociology, ethnography, statistics, optimization, and mathematics [3]. It encompasses the tools to formally represent, measure, model, and mine meaningful patterns from large-scale social media data. Social media mining is a rapidly growing new field. It is an interdisciplinary field at the crossroad of disparate disciplines deeply rooted in computer science and social sciences. There are an active community and a large body of literature about social media. The fast-growing interests and intensifying need to harness social media data require research and the development of tools for finding insights from big social media data [4].

1.2 Introduction to R

R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.

R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, …) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity. One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control [5].

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either on-screen or on hardcopy, and
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

2. TEXT MINING

Text mining is an emerging technology that attempts to extract meaningful information from unstructured textual data. Text mining is an extension of data mining to textual data. Study indicates that an estimated 80% of an organization's information is contained in text documents, such as emails, memos, customer correspondence, and reports [6].

Text mining is focused on finding useful models, trends, patterns, or rules from unstructured textual data such as text files, HTML files, chat messages and emails. As an automated technique, text mining can be used to "efficiently and systematically identify, extract, manage, integrate, and exploit knowledge from texts" [7]. Many researchers have successfully used text mining techniques to analyse large amounts of textual data in business, health science and educational domains used text mining techniques to extract metadata from documents in a digital library and to enrich documents by marking up appropriate items in the text [8].

Step 1: Text pre-processing is the process of cleaning the data. Need to create a error free data with proper format. The first step is annotation of target data, in the case of classification of problems and understanding the requirements.

Step 2: Text processing/Analysis is the process of deriving high amount of information from their respective sources. Where data can be finding out through manually like collecting likes/fans, followers, number of postings, comments or shares etc information from social media sites.

Step 3: Actionable Intelligence in applying different methods on the available datasets to find out patterns, issues or trends etc. Data visualization is the general method to describe the effort and results where people can easily understand and differentiate [9][10].

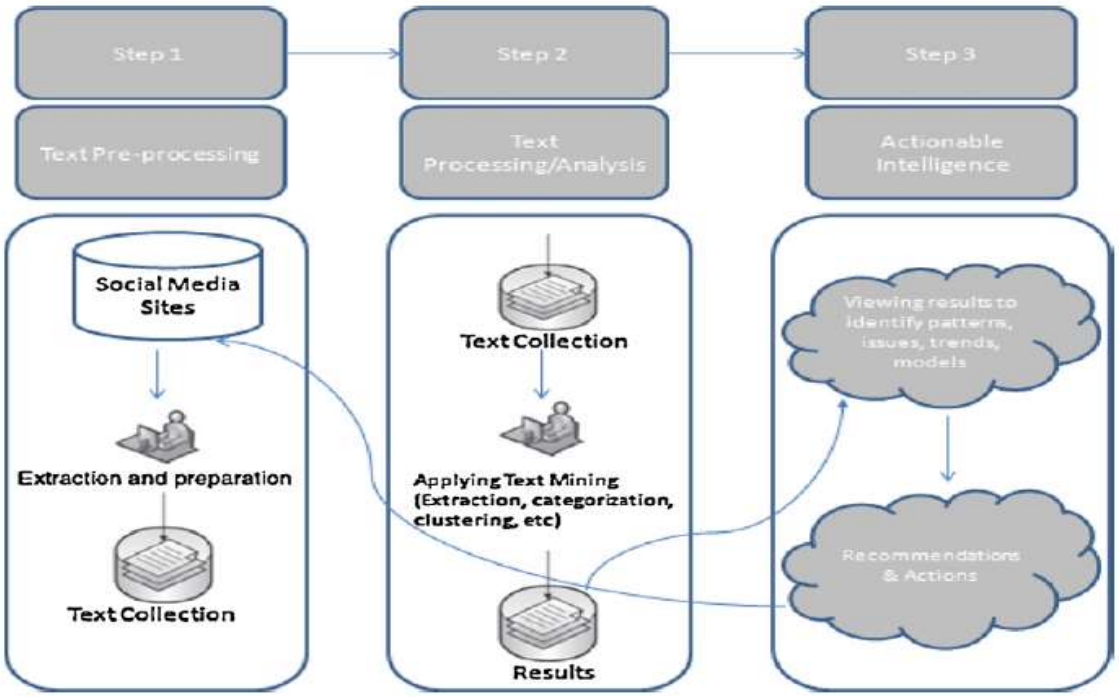

Fig -1 : Text mining process for social media content

3. CASE STUDY

User generated contents have been considered important in social media studies. In this study we have examined social media sites of three largest pizza chains and applied text mining methods to analyse unstructured contents on their different social media sites like Facebook and Twitter etc.

In initial phase we collected the data manually from their individual social media sites such as Followers, Page likes, Shares, Comments, Frequency of postings and number of Reponses etc. Following are the data of September 2017 in Table-1.

Table -1 : Figures of Social Media

Sr No	Industry Name	Facebook Page Likes	Twitter Followers
1	Pizza Hut	7,898,457	125,869
2	Domino's Pizza	7,657,956	168,248
3	Papa John's Pizza	4,251,965	86,254

In phase 2, we applied text mining to the text we collected in order to discover new knowledge and patterns. Social media data are usually large, noisy and unstructured. It would be tedious and time-consuming if we had to manually code a large amount of social media data. As there were no criteria available to compare the social media content of the three pizza chains, we decided to combine the textual data from the Big Three in order to discover main shared themes first. After themes were established, then we conducted query search based on each theme to examine and compare the detailed efforts of the Big Three.

The different Topics and their examples can be seen in the Table-2.

Table -1 : Examples related to the quality, ordering and delivering.

Topics	**Examples**
Customer sharing positive experiences with pizza quality	Wow ! Yummy Pizza
Customer sharing negative experiences with pizza quality	Topping is not good ! Quality is not soo good
Responses from customer service representatives	Sorry for the wrong topping ! Can you pls follow & DM your name, phone, email and store info? I'd like to help! Thanks, glad you like them.
Customer sharing positive experiences with online ordering and delivery	Having the iPhone app to order pizza is the greatest thing ever!! It's almost as awesome as delivery! You guys totally rock.
Customer sharing negative experiences with online ordering and delivery	Order did not have discount applied. I had to cancel the order for pizzas. Delivery so slow. Late almost 1 hour & 10 minutes.
Responses from customer service representatives	Thanks for the great feedback! We're so glad you like our iPhone app! Can you pls DM details and city where this happened? We take this seriously & I'd like to look into it. Thanks! We are very sorry for the experience you had! Can you pls DM your # so we can help make things right!

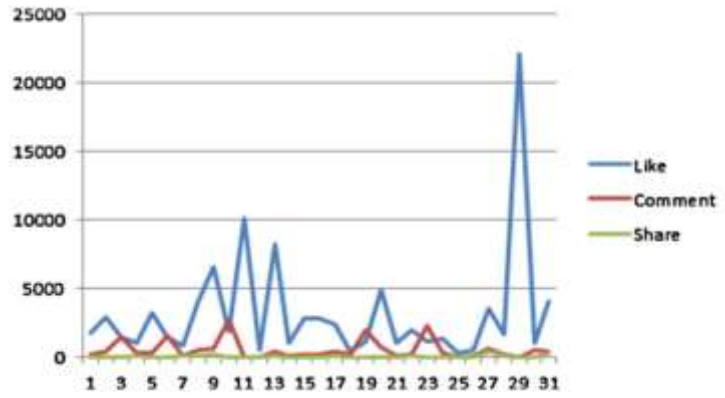

Fig -1: Pizza Hut's customer engagement trend

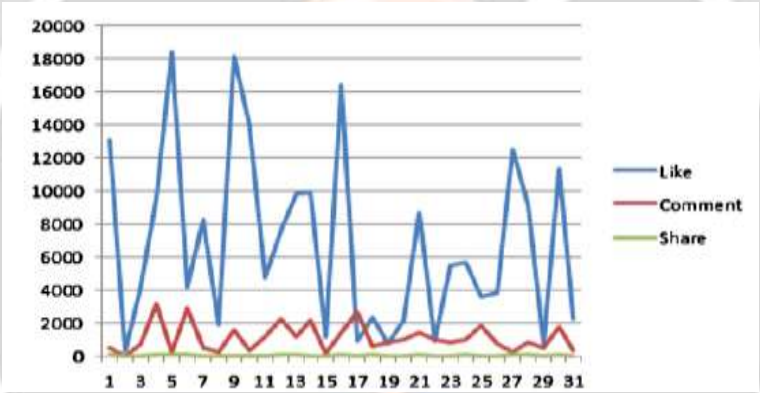

Fig -2: Domino's customer engagement trend

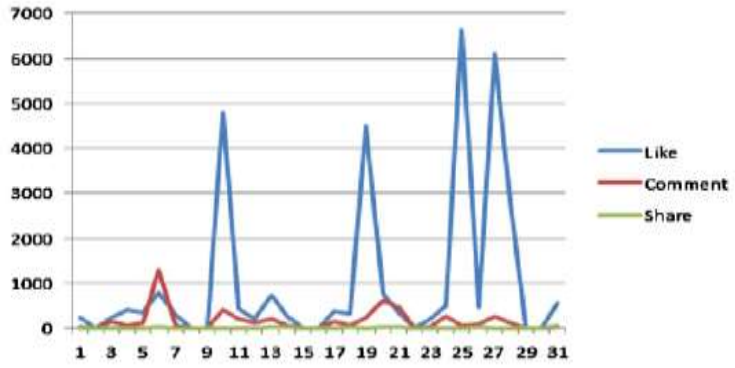

Fig -3: Papa John's customer engagement trend

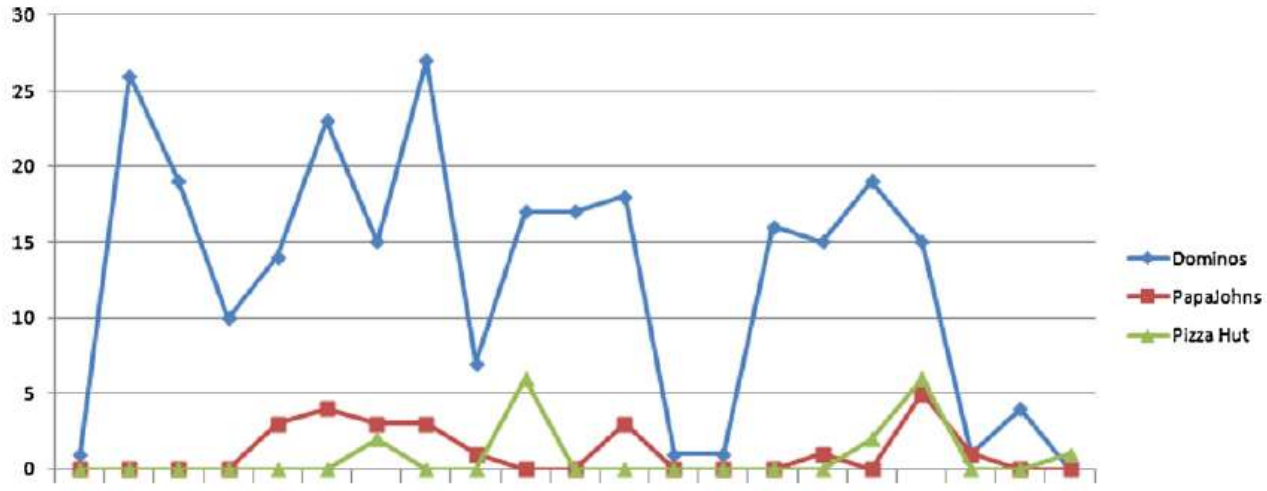

Fig -3: Trend of tweets numbers in September for the Big Three

4. CONCLUSIONS

As social media become topic of interest for many corporates. It is very important to understand the power of social media and data generated from social media can be used for decision making at industry level. Such analysis can be used for improving quality of products and services. Results from the text mining and social media competitive analysis show that these pizza chains actively engaged their customers in social media such as Twitter and Facebook. They used the social media not only to promote their services, but also to bond with their customers. Findings from this study suggest that social media plays an important role in sustaining a positive relationship with customers.

5. REFERENCES

[1]. Micheline Kamber. "Data Mining: Concepts and Techniques Second Edition Jiawei Han. University of Illinois at Urbana-Champaign."

[2]. Reza Zafarani, Mohammad Ali Abbasi, Huan Liu. "*Social Media Mining An Introduction*" Cambridge University."

[3]. Prasad V. , Dr. T. Nalini. "Social Network Analytics Data Mining Applied To Twitter Posts"

[4]. Fuller, C., Biros, D., & Delen, D. (2011). "An investigation of data and text mining methods for real world deception detection. Expert Systems with Applications."

[5]. Rubathee Nadaraja, Dr. Rashad Yazdanifard."Social media marketing: advantages and disadvantages.", Center of Southern New Hampshire University (SNHU) Of Help College of Arts and Technology, Kuala Lumpur, Malaysia

[6]. Wu Hea, Shenghua Zhab, Ling Li. "Social media competitive analysis and text mining: A case study in the pizza industry".

[7]. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135

[8]. Lin, F. R., Hsieh, L. S., & Chuang, F. T. (2009). Discovering genres of online discussion threads via text mining. Computers & Education, 52(2), 481–495.

[9]. Abdous, M., He, W., & Yen, C. J. (2012). Using data mining for predicting relationships between online question theme and final grade. Educational Technology & Society, 15(3), 77–88.

[10]. Anderson, M., Sims, D., Price, J., & Brusa, J. (2011). Turning like to buy social media emerges as a commerce channel. Retrieved from *www.booz.com/media/ uploads/BaC-Turning Like to Buy.pdf Accessed 07.06.2012.*