Social Media Mining and Sentiment Analysis for Brand Management

Umman Tugba Gürsoy,  
Istanbul University,  
Faculty of Business Administration, Istanbul, Turkey.  
E-mail: tugbasim@istanbul.edu.tr

Diren Bulut,  
Istanbul University,  
Faculty of Business Administration, Istanbul, Turkey.  
E-mail: dbulut@istanbul.edu.tr

Cemil Yiğit,  
Semanticum.co,  
School of Management, Istanbul, Turkey.  
E-mail: cyigit.99@gmail.com

Abstract

Data mining and Big data studies methods have attracted a great deal of attention in the information industry in recent years, due to the wide availability of huge amounts of data and the urgent need for turning such data into useful knowledge. Corporate firms want to benefit from big data studies more. Although it affects different company dynamics in various sectors, especially social media services have become very important for the marketing and CRM departments of companies. In this way, communication is always established with the customers and the use of Big data in these fields is seen as one of the most important steps of the companies in becoming a big brand. In this study, social media and digital data of the 3 major firms operating in the construction, technology and food industry in Turkey were analyzed. The data was obtained with the help of API and Web Crawler.

Key Words: Social Media Mining, Sentiment Analysis, Data Mining, Big Data

JEL Classification: C80, M30
1. Introduction

Data mining is the process of automatically discovering useful information in large repositories. Data mining techniques are deployed to scour large databases in order to find novel and useful patterns that might otherwise remain unknown. They also provide capabilities to predict the outcome of a future observation (Tan, Steinbach, Kumar, 2006).

Traditional data mining uses structured data stored in relational tables, spreadsheets, or flat files in the tabular form. With the grow of the web and text documents, Web mining and Text mining are becoming increasingly important and popular.

The Web has impacted on almost every aspect of people’s lives. It is the biggest and most widely known information source that is easily accessible and searchable. It consists of billions of interconnected documents which are authored by millions of people. Since its inception, the web has dramatically changed people’s information seeking behavior. Not only can we find needed information on the web, but we can also easily share our information and knowledge with others.

The web has also become an important channel for conducting businesses. People can buy almost anything from online stores without needing to go to a physical shop. The web also provides convenient means for people to communicate with each other, to express their views and opinions on anything, and to discuss with people from anywhere in the world.

2. Literature Review

The proliferation of Big Data & Analytics in recent years has compelled marketing practitioners to search for new methods when faced with assessing brand performance during brand equity appraisal. One of the challenges of current practices is that these methods rely heavily on traditional data collection and analysis methods such as questionnaires, and face to face or telephone interviews, which have a significant time lag. In this paper, the authors (Pournarakis, Sotiropoulos and Giaglis, 2017) introduce a computational model that combines topic and sentiment classification to elicit influential subjects from consumer perceptions in social media. Their model devises a novel genetic algorithm to improve clustering of tweets in semantically coherent groups, which act as an essential prerequisite when searching for prevailing topics and sentiment in big pools of data. To illustrate the validity of their model, they apply it to the Uber transportation network, from data collected through Twitter. The results obtained present consumer perceptions and produce insights for two fundamental brand equity dimensions: brand awareness and brand meaning.

Social media has generated a wealth of data. Billions of people tweet, sharing, post, and discuss every day. Due to this increased activity, social media platforms provide new opportunities for research about human behavior, information diffusion, and influence propagation at a scale that is otherwise impossible. Social media data is a new treasure trove.
for data mining and predictive analytics. Since social media data differs from conventional data, it is imperative to study its unique characteristics. This paper (Morstatter and Liu, 2017) investigates data collection bias associated with social media. In particular, the authors propose computational methods to assess if there is bias due to the way a social media site makes its data available, to detect bias from data samples without access to the full data, and to mitigate bias by designing data collection strategies that maximize coverage to minimize bias. They also present a new kind of data bias stemming from API attacks with both algorithms, data, and validation results. This work demonstrates how some characteristics of social media data can be extensively studied and verified and how corresponding intervention mechanisms can be designed to overcome negative effects. The methods and findings of this work could be helpful in studying different characteristics of social media data.

Kang, Wang, Zhang and Zhou’s work (2017) investigates the public’s opinions on a new school meals policy for childhood obesity prevention, discovers aspects concerning those opinions, and identifies possible gender and regional differences in the U.S. They collected 14,317 relevant tweets from 11,715 users since the national policy enactment on Feb 9, 2010 through Dec 31, 2015. They applied opinion mining techniques to classify tweets into positive, negative, and neutral categories, and conducted content analysis to gain insights into aspects of opinions in terms of target, holder, source, and function. The findings discovered the public’s opinions for policy improvement, contributed to the evidence base of health benefits for policy promotion and community collaboration, and revealed interesting gender and regional differences in the opinions. The social media analytics offers significant methodological implications for discovering the public opinions on food policies.

Thomaz’s paper (2016) proposes a social media content mining framework that consists of seven phases. The framework was tested empirically during the FIFA World Cup 2014 at Curitiba (Brazil) as one of the main host city destinations. The research focused on the mining of Twitter content with tourist services ontology (hospitality, food and beverages, and transportation). In total, 58,686 valid messages were collected, analyzed, and associated with an application ontology. Content analysis demonstrated an accurate real-time reflection of tourism services. The framework is effective to collect relevant content and identify popular topics in social media toward strategic and operational tourism management.

Today, the use of social networks is growing ceaselessly and rapidly. More alarming is the fact that these networks have become a substantial pool for unstructured data that belong to a host of domains, including business, governments and health. The increasing reliance on social networks calls for data mining techniques that is likely to facilitate reforming the unstructured data and place them within a systematic pattern. The goal of the paper (Injadat, Salo and Nassif, 2016) is to analyze the data mining techniques that were utilized by social
media networks between 2003 and 2015. They suggest that more research be conducted by both the academia and the industry since the studies done so far are not sufficiently exhaustive of data mining techniques.

Sun, Lanchanski and Fabozzi (2016) investigate the potential use of textual information from user-generated microblogs to predict the stock market. Utilizing the latent space model proposed by Wong et al., they correlate the movements of both stock prices and social media content. This study differs from models in prior studies in two significant ways: it leverages market information contained in high-volume social media data rather than news articles and it does not evaluate sentiment. They test this model on data spanning from 2011 to 2015 on a majority of stocks listed in the S&P 500 Index and find that their model outperforms a baseline regression. They conclude by providing a trading strategy that produces an attractive annual return and Sharpe ratio.

Twitter is one of the most widely used social media micro blogging sites. Mining user opinions from social media data is not a straight forward task; it can be accomplished in different ways. In this paper (Younis, 2015) an open source approach is presented, throughout which, twitter Microblogs data has been collected, pre-processed, analyzed and visualized using open source tools to perform text mining and sentiment analysis for analyzing user contributed online reviews about two giant retail stores in the UK namely Tesco and Asda stores over Christmas period 2014. Collecting customer opinions can be expensive and time consuming task using conventional methods such as surveys. The sentiment analysis of the customer opinions makes it easier for businesses to understand their competitive value in a changing market and to understand their customer views about their products and services, which also provide an insight into future marketing strategies and decision making policies.

Social media have been adopted by many businesses. More and more 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, He, Zha and Li’s paper [2013] describes an in-depth case study which applies text mining to analyze 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.

Blogs and social networks have recently become a valuable resource for mining sentiments in fields as diverse as customer relationship management, public opinion tracking and text filtering. In fact knowledge obtained from social networks such as Twitter and Facebook has been shown to be extremely valuable to marketing research companies, public opinion organizations and other text mining entities. However, Web texts have been classified as noisy as they represent considerable problems both at the lexical and the syntactic levels. In this paper, (Mostafa, 2013) used a random sample of 3516 tweets to evaluate consumers’ sentiment towards well-known brands such as Nokia, T-Mobile, IBM, KLM and DHL. He used an expert-predefined lexicon including around 6800 seed adjectives with known orientation to conduct the analysis. The results indicate a generally positive consumer sentiment towards several famous brands. By using both a qualitative and quantitative methodology to analyze brands’ tweets, this study adds breadth and depth to the debate over attitudes towards cosmopolitan brands.

Twitter messages are increasingly used to determine consumer sentiment towards a brand. The existing literature on Twitter sentiment analysis uses various feature sets and methods, many of which are adapted from more traditional text classification problems. In this research (Ghiassi, Skinner and Zimbra, 2013), the authors introduce an approach to supervised feature reduction using n-grams and statistical analysis to develop a Twitter-specific lexicon for sentiment analysis. They augment this reduced Twitter-specific lexicon with brand-specific terms for brand-related tweets. They show that the reduced lexicon set, while significantly smaller (only 187 features), reduces modeling complexity, maintains a high degree of coverage over their Twitter corpus, and yields improved sentiment classification accuracy. To demonstrate the effectiveness of the devised Twitter-specific lexicon compared to a traditional sentiment lexicon, they develop comparable sentiment classification models using SVM. They show that the Twitter-specific lexicon is significantly more effective in terms of classification recall and accuracy metrics. They then develop sentiment classification models using the Twitter-specific lexicon and the DAN2 machine learning approach, which has demonstrated success in other text classification problems. They show that DAN2 produces more accurate sentiment classification results than SVM while using the same Twitter-specific lexicon.

The Web holds valuable, vast, and unstructured information about public opinion. In this paper (Cambria, Schuller, and Havasi, 2013) the history, current use, and future of opinion mining and sentiment analysis are discussed, along with relevant techniques and tools.

www.globalbizresearch.org
By 2011, approximately 83% of Fortune 500 companies were using some form of social media to connect with consumers. Furthermore, surveys suggest that consumers are increasingly relying on social media to learn about unfamiliar brands. However, best practices regarding the use of social media to bolster brand evaluations in such situations remain undefined. This research (Naylor, Lamberton and West, 2012) focuses on one practice in this domain: the decision to hide or reveal the demographic characteristics of a brand's online supporters. The results from four studies indicate that even when the presence of these supporters is only passively experienced and virtual (a situation the authors term “mere virtual presence”), their demographic characteristics can influence a target consumer's brand evaluations and purchase intentions. The findings suggest a framework for brand managers to use when deciding whether to reveal the identities of their online supporters or to retain ambiguity according to the composition of existing supporters relative to targeted new supporters and whether the brand is likely to be evaluated singly or in combination with competing brands.

Social networks have changed the way information is delivered to the customers, shifting from traditional one-to-many to one-to-one communication. Opinion mining and sentiment analysis offer the possibility to understand the user-generated comments and explain how a certain product or a brand is perceived. Classification of different types of content is the first step towards understanding the conversation on the social media platforms. In this study (Cvijickj and Michahelles, 2011) analyses the content shared on Facebook in terms of topics, categories and shared sentiment for the domain of a sponsored Facebook brand page. The results indicate that Product, Sales and Brand are the three most discussed topics, while Requests and Suggestions, Expressing Affect and Sharing are the most common intentions for participation. The authors discuss the implications of our findings for social media marketing and opinion mining.

Hotel companies are struggling to keep up with the rapid consumer adoption of social media. Although many companies have begun to develop social media programs, the industry has yet to fully explore the potential of this emerging data and communication resource. The revenue management department, as it evolves from tactical inventory management to a more expansive role across the organization, is poised to be an early adopter of the opportunities afforded by social media. In this paper (Noone, McGuire and Rohlfis, 2011) the authors propose a framework for evaluating social media-related revenue management opportunities, discuss the issues associated with leveraging these opportunities and propose a roadmap for future research in this area.

3. Methodology

3.1 Social Media Mining and Sentiment Analysis
The web both contains a huge amount of information in structured and unstructured texts. Analyzing unstructured texts is of great importance and perhaps even more important than extracting structured data because of the sheer volume of valuable information of almost any imaginable types contained in them.

Businesses always want to find public or consumer opinions on their products and services. Potential customers also want to know the opinions of existing users before they use a service or purchase a product. Moreover, opinion mining also known as sentiment analysis, can also provide valuable information for placing advertisements in web pages. If in a page people express positive opinions or sentiments on a product, it may be a good idea to place an ad of the product. However, if people express negative opinions about the product, it is probably not wise to place an ad of the product. A better idea may be to place an ad of a competitor’s product.

Sentiment analysis is one of the approach that is used to analyze positive, negative and neutral opinion of people about specific brand or service.

Mining opinions on the web is not only technically challenging because of the need for natural language processing, but also very useful in practice.

The web has dramatically changed the way that people express their opinions. They can now post reviews of products at merchant sites and express their views on almost anything in Internet forums, discussion groups, blogs, Twitter, Facebook, Foursquare, Instagram, etc. This online word-of-mouth behavior represents new and measurable sources of information with many practical applications (Liu, 2008). Because of these features new techniques are needed and Social media mining has become popular.

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. It encompasses the tools to formally represent, measure, model, and mine meaningful patterns from large-scale social media data (Zafarani, Abbasi, Liu, 2014).

Twitter and Facebook are two of the todays the most known applications:

Twitter is a reach source of social data that is a great starting point for social web mining because of its inherent openness for public consumption, clean and well-documented API, rich developer tooling, and broad appeal to users from every walk of life. Twitter data is particularly interesting because tweets happen at the “speed of thought” and are available for consumption as they happen in near real time, represent the broadest cross-section of society
at an international level, and are so inherently multifaceted. Tweets and Twitter’s “following” mechanism link people in a variety of ways, ranging from short but often meaningful conversational dialogues to interest graphs that connect people and the things they care about (Russell, 2013).

Facebook is arguably the heart of the social web and is somewhat of an all-in-one wonder, given that more than half of its 1 billion users are active each day updating statuses, posting photos, exchanging messages, chatting in real time, checking in to physical locales, playing games, shopping, and just about anything else you can imagine. From a social web mining standpoint, the wealth of data that Facebook stores about individuals, groups, and products is quite exciting, because Facebook’s clean API presents incredible opportunities to synthesize it into information (the world’s most precious commodity), and glean valuable insights (Russell, 2013).

Companies have millions of tweets about their brands, thousands of Facebook “likes”, hundreds of thousands of check-ins on foursquare. Pinterest and Instagram are adding even more to social media data deluge. (Scarfi, 2012). Companies can manage their brand awareness and brand loyalty through social media. Positive, negative and neutral comments of people are core of social media analysis.

3.2 Data

The analysis focuses on the three industry-leading companies in different sectors: construction, food, and technology. The data were gathered from social media for the entire month of June 2016. Data were acquired three ways: from the Application Programming Interface (API), using HTML, and tracing mobile applications. Twitter API allows us to search Twitter. HTML tags provide context information that may be very useful in keyword searches. Naïve-Bayes, Support Vector Machine, and Neural Network Algorithms were used to automatically assess sentiment as positive, neutral, or negative. The results are shows in brand reports and infographics. Company names cannot be mentioned due to the privacy policy.

3.2.1 Construction Industry

In this study, the data of a large construction company in Turkey were analyzed. The sentiment distribution can be seen in Figure 1. There are totally 14,582 comments: 8878 of them are positive, 5650 are neutral and only 54 are negative. The gender distribution was 62% women and 38% men.
According to Figure 2 there were 10,026 comments made on Facebook, 1423 on Twitter, and 38 on Google.

The comment frequencies for each social media platform are in Figure 3. Comments were mostly made in Facebook (69%), followed by Instagram (21%) and Twitter (10%).

Most of the comments about the construction firm are neutral. On Facebook, 97% of comments are neutral and on Twitter, 96% are neutral. However, on Instagram, 87% of comments are positive.
The number of people on social media can be seen in Figure 5.

**Figure 4: The Results of Sentiment Analysis**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Positive</th>
<th>Neutral</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>3%</td>
<td>97%</td>
<td>0%</td>
</tr>
<tr>
<td>Twitter</td>
<td>2%</td>
<td>96%</td>
<td>2%</td>
</tr>
<tr>
<td>Instagram</td>
<td>87%</td>
<td>12%</td>
<td>1%</td>
</tr>
</tbody>
</table>

Table 1 shows the distribution of sentiment analysis by day of the week. Monday and Friday were the days with the most comments.

**Table 1: Distribution of Sentiment Analysis by Day of the Week**
In terms of location, the firm is mostly talked about in Beyoğlu (20%), followed by Pendik (15%) and then Tuzla (10%). There are ongoing construction projects in these locations.

![Figure 6: Locations]

The distribution of dates are in Figure 7. Most comments were made during June 20–24.

![Figure 7: Distribution of Dates]

### 3.2.2 Food Industry

The largest and best-known firm in the Turkish food industry was chosen for social media mining. The sentiment distribution is in Table 2. Most of the comments are positive and were made during June 1–10. 61% of the content was created by men.

<table>
<thead>
<tr>
<th>Sentiment</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>28,377</td>
</tr>
<tr>
<td>Neutral</td>
<td>14,340</td>
</tr>
<tr>
<td>Negative</td>
<td>2549</td>
</tr>
<tr>
<td>Total</td>
<td>45,266</td>
</tr>
</tbody>
</table>

The top three cities in Turkey were Istanbul (29%), Ankara (13%), and İzmir (13%). The remaining 47% of comments are from other cities.

The distribution of comments made on different social media sources can be seen in Table 3.

<table>
<thead>
<tr>
<th>Social Media Source</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instagram</td>
<td>19,767</td>
</tr>
<tr>
<td>Twitter</td>
<td>13,658</td>
</tr>
<tr>
<td>Facebook</td>
<td>11,731</td>
</tr>
<tr>
<td>Google+</td>
<td>95</td>
</tr>
<tr>
<td>Vine</td>
<td>10</td>
</tr>
</tbody>
</table>
The majority of comments were made on Instagram (44%), followed by Twitter (30%) and Facebook (26%).

Figure 8: Pie Chart for Social Media Sources

<table>
<thead>
<tr>
<th>Social Media</th>
<th>Positive</th>
<th>Neutral</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>85</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Twitter</td>
<td>29</td>
<td>62</td>
<td>9</td>
</tr>
<tr>
<td>Instagram</td>
<td>72</td>
<td>27</td>
<td>1</td>
</tr>
</tbody>
</table>

The results of sentiment analysis are below.

Table 4: The Results of Sentiment Analysis

Data visualizations such as charts, graphs, and infographics give businesses a valuable way to communicate important information at a glance. If the data is text-based and the analyst wants a stunning visualization to highlight important data points, using a word cloud can make dull data sizzle and immediately convey crucial information. Word clouds work in a simple way: the more often a word appears in the source, the bigger and bolder it appears in the word cloud. A sample word cloud can be seen in Figure 9.

Figure 9: Word Cloud

Most of the comments were made on June 9–10. Positive comments were mostly made on Wednesdays, Mondays, and Tuesdays.
3.2.3 Technology industry

For the technology industry, data for a technology firm were analyzed. The sentiment distribution is in Table 5. More comments were made in this sector than the others.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>68,206</td>
</tr>
<tr>
<td>Neutral</td>
<td>39,731</td>
</tr>
<tr>
<td>Negative</td>
<td>5,381</td>
</tr>
<tr>
<td>Total</td>
<td>113,318</td>
</tr>
</tbody>
</table>

The results of the sentiment analysis can be seen in Figure 11.

The comments were mostly made during June 11–21 with the number of 43,535.
According to Figure 12 shows three cities in Turkey with the most comments about the brand. The percentages can be seen in the infographic.

**Figure 12: Location infographic**

4. Summary

In this paper, social media mining and sentiment analysis were used to analyze social media data for three industry-leading companies in Turkey. Companies in the construction, food, and technology sectors were chosen for analysis. Brand reports and infographics were used to describe the data, which were gathered from social media during the month of June 2016. Graphics, pie charts, and word clouds were used to show distributions of sentiment, location, gender, and dates.

5. Conclusions and Recommendations

Businesses always want to know public or consumer opinions about their products and services. Potential customers also want to know the opinions of previous customers users before they use a service or purchase a product. Opinion and sentiment analyses provide valuable information for placing advertisements on web pages. Future social media mining projects could focus on crises or web advertisements.

**Acknowledgement:**

This work is supported by research fund of Istanbul University (BAP) with the project number 53625.

**References**


Tan, P.N., Steinbach, M. and V. Kumar, 2006, Introduction to Data Mining, Pearson, 3.

