Received February 12, 2019, accepted March 9, 2019, date of publication March 15, 2019, date of current version April 3, 2019.
Digital Object Identifier 10.1109/ACCESS.2019.2905301

Comparing a Traditional Approach for Financial Brand Communication Analysis With a Big Data Analytics Technique

JOSE RAMON SAURA, BEATRIZ RODRÍGUEZ HERRÁEZ, AND ANA REYES-MENENDEZ
Department of Business Economics, Rey Juan Carlos University, 28032 Madrid, Spain
Corresponding author: Ana Reyes-Menendez (ana.reyes@urjc.es)

ABSTRACT Although large amounts of data are now available to companies, mere possession of these data is not sufficient, and for better business decisions, it is necessary to perform thorough data analysis. Nowadays, social networks services (SNS) have become important data sources. The rapid growth of SNS has led to their wide use in various research trends in social sciences. In this paper, we aim to enhance the current understanding of the possibilities offered by social data for brand communication analysis in the financial sector. To this end, a traditional methodology and a digital methodology are used to investigate the brand image of the financial entities. The traditional methodology is the Periodic Evaluation of the Image (PEI). The digital methodology is sentiment analysis, a machine learning technique for big data analytics in social sciences using an algorithm developed in Python. The data are analyzed using both methodologies, and then, their results are compared. The findings suggest that while the results obtained using the method based on big data are consistent with the results obtained with the traditional methodology, the former method allows for easier and faster data analysis. The limitations of this paper relate to the size of the sample, the studied sector, and the scope of the reviewed literature.

INDEX TERMS Big data applications, machine learning, sentiment analysis, social network services, Twitter.

I. INTRODUCTION

Recent years have witnessed a rapid growth in the amount of data available about companies on the Internet. Social Network Services (SNS), an important source of such data [1], have come to be widely used in the emerging research trends in social sciences [2]. Likewise, technology, which favors the welfare of human beings and helps to improve the environment, has played an essential role in the development of our society [2]. Furthermore, due to the increase of access to the Internet worldwide, we can speak about a democratization of the access to information [3].

In 2015, there were 3,000 million Internet users. In Europe, more than 80% of the inhabitants aged between 16 and 74 years old accessed the Internet for different purposes [1], [4]. This access was made using different devices, although mobile phones were the preferred means, followed by laptops, desktops and, finally, tablets [3]. At present, smartphone sales are the highest worldwide and are expected to exceed those of personal computers (PC) soon [2].

Mobile phones are available to more than 90% of the global population, and mobile phone adoption rate advances at a great speed. It is expected that, in 2019, the penetration of the mobile Internet will be 71%, while the use per device will reach 71% of the population [1], [3].

This access to the smartphones, which now perform the previously reserved for computers and laptops, is modifying the way people communicate today [3], [4]. Some of the tasks performed nowadays with mobile phones include online purchases, getting financial services, or getting access to information and news.

These changes should be taken into account in marketing research [4]. In general, users use traditional and digital media, and companies interact with their consumers through both traditional and digital media channels. However, new technologies open up new horizons for companies in terms of the development of their communication with customers and the creation of their brand image. In this context, it is necessary enhance our current understanding of and knowledge

The associate editor coordinating the review of this manuscript and approving it for publication was Pasquale De Meo.

about the relationship between traditional and digital media and how the two influence consumer perceptions.

To this end, it is necessary to use new techniques such as those based on Big Data analytics. The ultimate goal of Big Data analytics is to find trends, correlations, patterns, insights, or consumer preferences to improve business decision-making process [5]. Big Data analytics is the process of analyzing large amounts of information organized in a structured or unstructured way. In order to perform this analysis, it is first necessary to search for a database, organize it with data text mining techniques, and, finally, proceed with its analysis in order to be able to base decision making on data. However, this three-step process has to face several challenges, particularly in the initial phase of obtaining the data, since the sources tend to be diverse, which complicates obtaining and later analyzing the data [6].

Of note, considering the present-day interaction of the traditional channel and the digital channel, it has to be investigated whether the same results about the reputation image of a company can be obtained using different methodologies [7]. Therefore, the main aim of the present study is to enhance current understanding of the possibilities that social data offer for the brand communication analysis in the financial sector.

The remainder of this paper is structured as follows. In Section II, we discuss the financial sector in Spain. Section II reviews relevant literature. In Section IV, we outline and discuss the two methodologies used in the present study. The traditional methodology used is the Periodic Evaluation of the Image (PEI), and the digital methodology is a Sentiment Analysis, a machine learning technique for Big Data analytics in social sciences using an algorithm developed in Python. In Section V, we present the results. This is followed by Section IV where the findings are discussed with particular focus on the degree of consistency between the two methodologies. Finally, conclusions are drawn in Section VII.

II. THE FINANCIAL SECTOR IN SPAIN

In recent years, there has been a very intense change in the financial sector worldwide. In 2007, an international financial crisis threatened global stability [11]. The crisis was caused by the fall of Lehman Brothers that led to a loss of confidence in the banking system which, in turn, resulted in that some financial entities had to be rescued by their governments.

In parallel to the 2008 crisis, the world's economy entered a phase of recession that caused a weakening of credit operations for customers, which resulted in lower consumption and spending capacity and deepened the recession. In Spain, the real estate bubble had an adverse effect on the financial sector, because it uncovered the situation of savings banks, i.e. entities of a charitable nature, earmarking their profits for social projects.

Consequently, along with limitations in legal and corporate governance structure, savings banks suffered a great exposure to the real estate sector. This hindered the solvency ratios and the survival of these entities. One of the initial solutions that were attempted then was the merger of the savings banks. However, in 2012, Spanish banks had to ask for investment assistance from the European Union. To perform the reorganization of the sector, public capital had to be injected, which entailed the control of part of the banking system by the government of Spain. This process of reorganization led to the disappearance of many savings banks, so that their number decreased from 45 to 8, and to transforming them into financial entities. At present, there are five banking groups that accumulate almost 80% of the assets in Spain: Santander, BBVA, CaixaBank, Bankia, and Sabadell.

Another fundamental change in the sector has been the expansion of Internet banking that has evolved by leaps and bounds as a result of the normalization of the use of new technologies. In this context, it is interesting to explore whether and if, so, how the public image of savings banks in Spain has been damaged by the 2008 economic crisis. To explore this issue, in the present study, we use two methodologies—a traditional and a digital one applied to the user generated content (UGC) in Twitter [4], and then compare their outcomes.

III. RELATED WORK

Public image is an important factor when deciding and selecting a product, a service, or a brand. In their decision- making process, consumers evaluate the attributes of different products, services, or brands. Given that, today, the media are changing, leading to a convergence of communicative interactions and on-line and off-line communication, it is necessary that the management of the brand image incorporates the new reality. According to Blanco-Calleja [8], brands are valuable resources for companies, as they can provide a long-term competitive advantage [9] that can be controlled by the company itself [10]. As a result, financial companies seek to develop their brands. One of the forms of such brand development is sponsorship [11]. For instance, Banco Santander stands out as an example of a successful brand that has extensively used sponsorship as a way to enhance the company's brand image [12].

The method of Periodic Evaluation of the Image (PEI) is developed with projective techniques and in-depth interviews and allows to obtain the key elements of the brand image and to establish the relations between the different attributes of the brand [13]. This periodic analysis of the image can be performed for different sectors, ranging from the financial sector [14] to the tourism sector [15]. For instance, using PEI that combines the benefits of quantitative and qualitative methodologies, Herraez Martínez *et al.* [14] analyzed the brand image of different financial organizations in Spain. The authors found that, while some financial entities had brand images that were perceived as very similar, others were perceived as very different.

Furthermore, Orihuela and Cambronerom [7] found that, with the introduction of SNS and new media, companies are no longer in full possession of the brands—rather, it is in the process of interaction with consumers that brand images are constructed. Accordingly, consumers become prosumers on the Internet.

TABLE 1. Previous research on brand image in the financial sector.

	Simoes et al. [18]	Acevedo et al. [17]	Bong-Gyu and Ho-Seok [6]	Mattar [19]	Blanco-Callejo [8]	Cerviño [11]	Küster et al. [12]	Kwon et al. [20]
Traditional brand image	-	-	-	-	✓	✓	✓	✓
Online brand image	✓	✓	✓	✓	-	-	-	✓

Of all online platforms that can be used to analyze company brands, Twitter is the most popular social platform [4], [17]. For instance, in a study on bank images based on the analysis of their Twitter accounts, Simoes *et al.* [18] developed and used a sentiment analysis methodology. Likewise, Mattar [19] also performed an analysis of SNS for the financial sector. Furthermore, Bong-Gyu and Ho-Seok [6] identified the brand image and purchase intentions of the consumers of an online bank called WeBank. To this end, the authors used factor analysis and multivariate analysis.

Considering that brand image is created using different channels, it appears to be necessary to use different methodologies to study it. This is exactly what Shahzad *et al.* [20] did in their research on the factors that influence buying behavior based on the brand image both online and off-line.

At the same time, several studies have investigated the consistency between brands and how they communicate their image in both on-line and off-line environments [21]–[23].

Table 1 summarizes previous research on brand image in the banking sector organized according to the approach that was used (traditional vs. online/digital).

IV. METHODOLOGY

A. THE PEI METHOD

The PEI is a research technique that has different applications in the field of marketing. Generally, the PEI is typically used to determine the image of a company or a brand. In the PEI method, the first step data collection. What are collected are perceptions that can be quantified and, therefore, can be expressed numerically. To obtain the data, a set of "stimuli" (such as brands, products, entities) is shown to each interviewee; these stimuli are used to project the opinions. Each stimulus is represented with the interviewer's name on a card.

In the second step, the three-in-three stimuli are shown, so that, at the end of the interview, all the three-in-three combinations that can be formed with the stimuli will have been shown. According to statistical laws, the number of possible combinations that can be made with three-in-three stimuli is defined by Eq. (1).

$$\mathrm{nCq} = \frac{n!}{q!(n-q)!} = \frac{npq}{q!} \tag{1}$$

where nCq = Combination of n elements taken from q to q
n = number of different elements
q = Quantity in which n objects combine
nPq = Permutations of n elements taken from q to q

For this research technique to be operative and the answers to be valid, the number of stimuli presented to the interviewees should five up to seven [23]. Using fewer than five stimuli would make the collected data sample too small. In contrast, more than seven stimuli are shown, the method is not operative, and the time needed for each interview would be very long.

Thus, in the PEI method, the interviewees are shown triplets of stimuli and are asked to group two of these cards based on their similarity on some important feature (which is not possessed by the third stimulus). The characteristics noted by the interviewees are then collected and grouped according to their meaning by proximity, until the list of attributes is reduced as much as possible. The characteristics are evaluated in terms of their valence (i.e. positive or negative).

With appropriate data processing, this method makes it possible to obtain different types of information, such as general attributes of the image, association between stimuli, relative image of the stimuli, individual image of each stimulus, integral image of each stimulus, and global positioning of the stimuli [17].

B. THE SENTIMENT ANALYSIS

Sentiment analysis is a methodology used to measure and identify feelings related to a specific event. This methodology has evolved in recent years and its major aim is to capture and categorize the feelings expressed in a given sample by dividing them into positive, negative, or neutral sentiments. In the present study, we performed a sentiment analysis with an algorithm developed in Python. The algorithm was accessed through the MonkeyLearn API libraries [24]. According to Saura *et al.* [4], this technique can be used in different ways. For instance, approximations can be made using special software to apply machine learning, artificial intelligence techniques, and hybrid models. Other available options include algorithm training with data mining techniques, which are processes used to improve the probability of success of an algorithm with machine learning based on the accuracy of the results [27], [28]. Supervised methods using the classification and categorization of key factors, such as Maximum Entropy (MaxEnt) and Support Vector Machines (SVMs) have also been widely used to perform social network analysis with machine learning using technological research methods to identify the important factors in different areas of research [14], [15].

As suggested in previous research, sentiment analysis can also be combined with other data extraction technologies based on Big Data or digital data management [e.g., 25]. In the present study, the development of the methodology is based on the study by Ekenga *et al.* [26], Simoes *et al.* [18], Palomino *et al.* [25], and Reyes-Menendez *et al.* [27].

In terms of data collection, a total of 7,598 tweets with the hashtags (#, see [4]) of the companies under study (#Santander, #BBVA, #LaCaixa, #ING, #Bankia and #Sabadell) were downloaded. After the filtering of the database, 198 tweets were eliminated, so a total of 7,398 tweets remain in the dataset. The data collection period was September 1-7, as, throughout this period, there were no major events that could have influenced the public images of the companies under study, and consequently, there was no impact on those brand images of SNS user opinions. Also, given that the Twitter API does not allow tweets to be downloaded after 7 days, the tweets were collected on September 7.

SUPPORT VECTOR MACHINE (SVM)

The next step was to train the SVM algorithm developed in Python. The SVM algorithms work with machine learning for Big Data analytics and are trained with data mining to increase the credibility of their results. In the present study, we connected to the public API of Twitter to download the tweets related to the profiles of the companies under study. Next, in order to analyze the comments of the users regarding the tweets made by these companies in the SNS Twitter, the SVM sentiment analysis algorithm was used.

First, the tweets with the aforementioned hashtags were downloaded, which was followed by the analysis of the sentiments expressed in the tweets. Before sentiment analysis, the collected data were filtered and cleaned using the following criteria:

•Retweets were removed

•URLs were removed

•The selected tweets had to contain a minimum of 80 characters (without the hashtag)

•The interaction with the company was eliminated from the tweets, e.g., @Bancosantander

•Special characters such as "", "=" or "!" were replaced by spaces.

Second, as in our previous studies (e.g., [26], [28]), once the database was cleaned, the sentiment analysis algorithm was applied separately to the tweets of each of the companies under study. Before classifying the tweets in different sentiments (negative, neutral, and positive), we trained the algorithm with data-mining processes so that to increase the probability of its success. The probability of success of the algorithms that work with a machine learning technique for Big Data analytics in social sciences is known as "accuracy". In this case, after training the sentiment analysis algorithm with 521 samples taken from Twitter, an average Accuracy percentage of 0.711 was achieved. This percentage is above the conventionally used threshold of > 0.600 [4], [27], [28]. In this way, we ensured that the algorithm and analysis of sentiment used was appropriate for the sector and object of study.

It should be noted, however, that a weakness of the sentiment analysis with a machine learning technique for Big Data analytics in social sciences is overlooking the context in which the tweet is made. Furthermore, sentiment analysis is less relevant for the analysis of content that contains sarcasm, irony, or metaphor. However, following Palomino *et al.* [25], Bennett *et al.* [29], and Reyes-Menendez *et al.* [27] we can confirm that the use of this methodology is totally valid.

V. ANALYSIS OF RESULTS

A. RESULTS OF THE PEI METHOD

Using the PEI method, first, the data were obtained. Thereafter, we obtained general attributes of the image, i.e. different attributes that the interviewees noted for each stimulus (i.e. each financial entity). The attributes with the highest frequency of appearance are the one that are most salient in the minds of the interviewees.

In our case, the number of perceived attributes was not very high (≤50). The distribution of these attributes was very uneven, with the first three attributes accumulating over 50% of all responses (see Table 2 for further detail).

TABLE 2. Attributes of the studied financial entities.

Characteristics	Frequency	Cumulative frequency
Give more confidence / give more security (+)	20,4	20,4
More accessibility / proximity (+)	15,4	35,8
More advertising (+)	13,4	49,2
Best online services (+)	9,3	58,5
Better resource management (+)	4,0	62,5
Personal attention / service (+)	4,0	66,5
Adapted to young people (+)	3,7	70,2
Less confidence (-)	3,5	73,7
Social projects / social work (+)	3,2	76,9
Larger variety of products (+)	2,5	79,4
Greater solvency (+)	2,2	81,6
Fewer commissions (+)	2,2	83,8
Custom products (+)	2,0	85,8
More innovative (+)	1,7	87,5
Not been intervened (+)	0,7	88,3
Other features	11,7	100,0

In the PEI methodology, the level of association is greater when more interviewees associate any pair of attributes a greater number of times, and vice versa. Among the analyzed entities, the two that were perceived as more equal among themselves according to the interviewees were Banco Santander and BBVA, with an association level

TABLE 3. Association between variables.

Entities	Frequency of association
Santander/BBVA	45,0
Santander/La Caixa	32,5
Santander/ING	31,7
Bankia/La Caixa	30,0
BBVA/ING	29,2
BBVA/Bankia	25,8
BBVA/La Caixa	24,2
Santander/Bankia	20,0
Sabadell/Bankia	20,0
Bankia/ING	18,3
Sabadell/La Caixa	15,8
La Caixa/ING	15,0
Sabadell/ING	10,8
BBVA/Sabadell	8,3
Santander/Sabadell	6,7

of 45.0 percent. On the contrary, the two entities perceived as the most different were Banco Santander and Banco Sabadell. Table 3 shows the level of association between the financial entities under study.

The relative image of the stimuli represents the perception that the interviewee has of each stimulus. If all the stimuli are perceived as identical, their value will correspond to the quotient of dividing one hundred among the total number of stimuli analyzed. When a stimulus has a value higher than this value, it means that, in that stimulus, the attribute is higher than the average. In our case, if we take, for instance, the attribute that is repeated more frequently, which corresponds to the ''give more confidence / give more security'', we can see how differently it is perceived for each entity.

The calculation of the theoretical average gives the value of 16.7 percent; therefore, as concerns our results, we can conclude that, except for Banco Santander and BBVA (with the values of 29.9 and 23.2 percent, respectively), the remaining entities have a lower perception than the theoretical average, and it includes the evaluation of all the attributes in relation to the theoretical value of the image, taking into account its positive or negative character.

The vertical position of each stimulus results from assigning the value zero to the stimuli with less evaluation and the one hundred to the stimulus of greater evaluation, while the other stimuli will have a value between both calculated proportionally to the score (see Table 4).

The individual image of each stimulus represents the perception that the interviewees have of each stimulus without considering the remaining brands. With respect to the graphics of the individual image of each entity, it should be mentioned that, since the concentration of the perception of the general attributes for the entities analyzed is high, the individual image of each entity is also high.

Finally, the global position of the stimuli represents the position of each stimulus in relation to the others vis-à-vis the set of attributes. Figure 1, which includes this positioning, shows he horizontal position of each stimulus.

TABLE 4. Perception of the attribute "give more confidence / give more security between the different entities.

Entities	Percentage
Banco Santander	29,9
BBVA	23,2
Banco Sabadell	6,7
Bankia	13,4
La Caixa	11,6
ING	15,2
Total	100,0

B. RESULTS OF SENTIMENT ANALYSIS

Sometimes, social data generated by users are difficult to obtain. In the present study, we collected the data using the mechanism based on machine learning to download and analyze tweets. As mentioned above, the process was three-fold. First, we extracted the data from the public Twitter API in which we consulted the name of the profile of the companies under study, followed by any user opinions, comments or tweets that mention to those companies [30], [31].

In order to have access to the Twitter API, we logged into Twitter Developer and created an application. The information regarding the application name, description, and contact data is filled in with respect to the web page accessed by the Developer [4]. After registration, a token linked to the key of the created Application was created. These data were collected for subsequent data extraction. Figure 2 shows the process used for the development of this approach. In addition, the number of tweets identified according to each sentiment (t) is shown in Figure 2.

An issue that usually arises is the accuracy of the analyzed data. Therefore, it is essential to incorporate mechanisms to identify possible inaccuracies in the analyzed data [4]. In our case, accuracy referred to the success of the automatically analyzed data (i.e. the probability of success obtained after algorithm training [4]). In order to increase the accuracy of the used sentiment analysis algorithm, we trained with the data-mining processes and subdivided the sample into positive, negative and neutral according to the sentiments about the financial companies' industry expressed in user comments. As indicated before, a total of 521 samples were processed using the data-mining mechanism [4].

The total number of tweets analyzed included only the interactions and comments made directly by users on the profiles of the financial institutions under study. As defined by Saura *et al.* [4], users' perceptions expressed on digital platforms and in online environments are collectively referred to as user generated content (UGC). Over the last decade,

UGC has been widely used to determine the key factors for any specific topic [4], [24], [38]. Table 5 shows the basic characteristics of the analyzed profiles of financial entities.

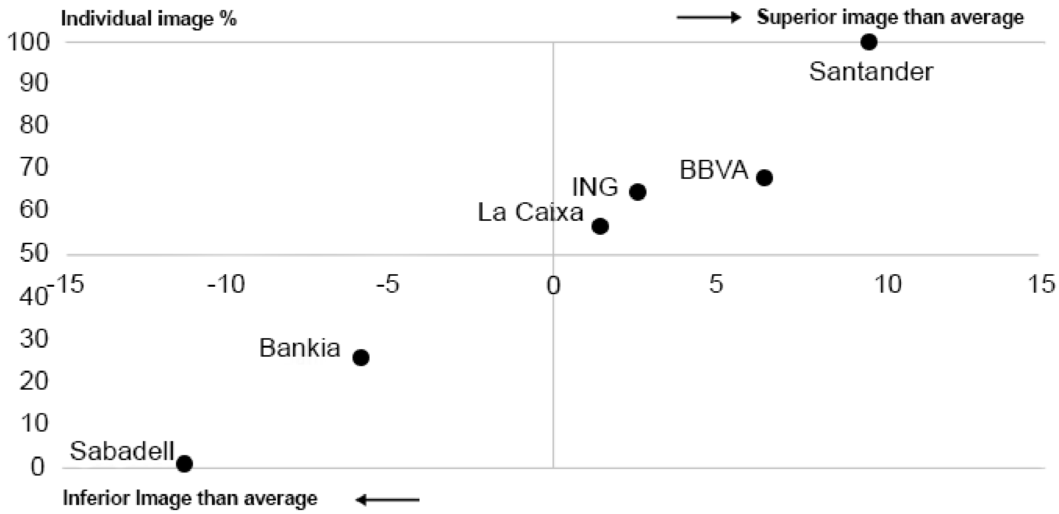


FIGURE 1. Global positioning of the entities under study.

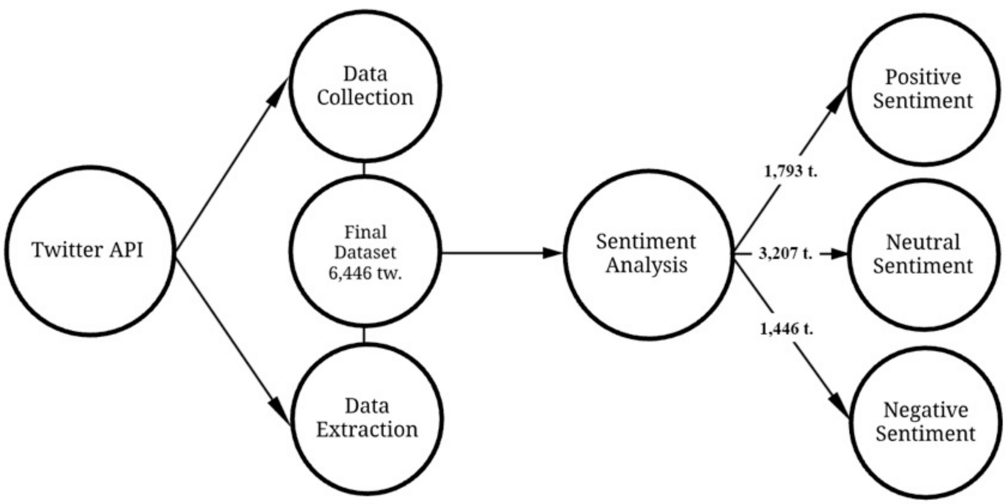


FIGURE 2. Data extraction process.

TABLE 5. Number of extracted tweets, profile name, and number of followers.

HashTags	Tweets	Profile	Followers
#Santander	2002	@bancosantander	57.8k
#BBVA	2406	@bbva	95k
#Caixabank	1253	@caixabank	41.8k
#ING	26	@ING_news	38.9k
#Bankia	955	@Bankia	25.1k
#Sabadell	756	@BancoSabadell	43k

Algorithm training was performed after identifying the contents exclusively related to the research topic, including also ironic and sarcastic comments included [17]. Throughout the entire process, the contents that were not related to the financial industry were discarded from the sample and training. The average accuracy for positive tweets was 0.679 percent, 0.701 percent for neutral tweets, and 0.702 percent for negative tweets (see Figure 3). The most commonly used measure to measure the accuracy of a sentiment analysis algorithm is Krippendorff's alpha. Krippendorff's alpha should be equal or above 0. 667.

Values above this threshold indicate that the algorithm has been trained a sufficient number of times and its predictive capacity is sufficiently high [17].

In the absence of knowledge of the risks of making false conclusions from unreliable data, social scientists rely on the data with reliabilities $\alpha \geq 0.800$; furthermore, the data with $0.800 > \alpha \geq 0.667$ are considered only to draw

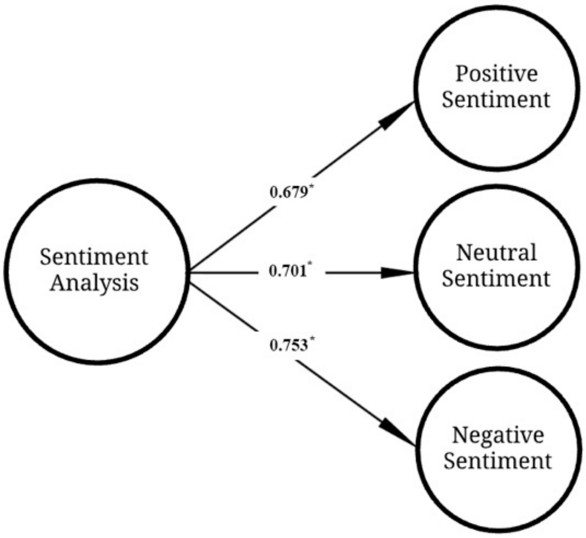


FIGURE 3. Sentiment Analysis conclusions' reliability of Krippendorff's alpha value.

tentative conclusions; finally, the data with agreement measures $\alpha < 0.667$ are discarded (see Table 6).

TABLE 6. Sentiment analysis conclusions' reliability (Krippendorff's alpha).

Conclusions reliability	Krippendorff's alpha value
High	$\alpha \geq 0.800$
Tentative	$\alpha \geq 0.667$
Low	$\alpha < 0.667$

The average accuracy across three types of sentiments was 0.711. The total of positive tweets was 1,793; furthermore, 3,207 tweets were neutral, and 1,446 tweets were negative.

In the second step, the sentiment analysis algorithm that classifies each of the tweets to negative, neutral, or positive feelings was applied. As indicated before, a total of 1,793 positive tweets, 3,207 neutral tweets and 1,446 negative tweets were obtained and further subdivided by financial entity and accuracy percentage (see Table 7).

TABLE 7. Sentiment of tweets according to company.

Entities	Positive	Neutral	Negative	Average Accuracy
Santander	543	703	160	0.689
BBVA	632	940	308	0.709
La Caixa	198	628	430	0.652
ING	3	17	6	0.812
Bankia	26	535	561	0.734
Sabadell	44	384	328	0.620

The average of the order percentage (P) of positive and negative tweets is shown in Table 7. There were on average 14.5% positive tweets, 48% neutral sentiment tweets, and 29% negative tweets. These results suggest that users more frequently express negative rather than positive opinions about financial companies on Twitter [4].

Furthermore, as concerns the brand image of the studied financial entities, the order of companies depending on the number of positive sentiment tweets they received was as follows: Santander (27), BBVA (26), La Caixa (15), ING (11), Sabadell (5), and Bankia (3).

Likewise, the order of companies depending on the number of negative sentiment tweets they received was as follows: Bankia (54), Sabadell (43), La Caixa (34), ING (23), BBVA (12), and Santander (8).

Neutral tweets were defined as contents that do not affect the reputation of the brand (neither positively nor negatively).

VI. DISCUSSION

In the present study, we focused on investigating consistency and complementarity of two research methodologies and their results. The used methodologies were the PEI method and sentiment analysis. Both methods have been widely used in the field of marketing to study the brand image from the traditional and digital perspective.

In our results, an interesting finding that we obtained using the PEI method was the importance that respondents attributed to confidence / security. This importance of confidence and security can be explained by the economic situation observed in recent years in Spain—specifically, throughout the period of insecurity and distrust in the financial industry. Furthermore, our finding that the two brands that were identified as more similar were Banco Santander and BBVA and that the two received the highest number of positive comments (the average accuracy: 0.679) could be explained by the fact that, in recent years, these two entities have become known for their solvency and their lack of involvement into all types of conflicts or bad business practices, which has allowed them to position themselves as two solvent companies in the financial market. By contrast, the financial entities with the worst negative results were Bankia and Sabadell (the average accuracy: 0.702). As in the case outlined above, these results are directly linked to the reputation and corporate identity policies maintained by financial entities in the digital environment.

Furthermore, we also observed a strong coherence between the recent events in Spain and the perception of the brand images of the financial entities by our interviewees. More generally, our results highlight that it is possible to use data from SNS such as Twitter to measure use feelings towards brand images of financial entities [15], [36]. In recent years, Twitter has emerged in the new digital paradigm as a SNS where users make comments and show their opinions on specific issues and which are linked around #Hashtags. In the present study, we have demonstrated that this new bidirectional paradigm of digital communication of the 21th century

TABLE 8. Average sentiment of tweets according to company.

		Positive		Neutral		Negative	
Entities	**Total**	**Tweets**	**Percentage**	**Tweets**	**Percentage**	**Tweets**	**Percentage**
Santander	2002	543	27	703	35	160	8
BBVA	2406	632	26	940	39	308	12
La Caixa	1253	198	15	628	50	430	34
ING	26	3	11	17	65	6	23
Sabadell	756	44	5	384	48	328	43
Bankia	955	26	3	535	56	561	58

can be meaningfully used for the analysis of the reputation of financial companies [4], [32], [33].

VII. CONCLUSIONS

The growing number of social platforms and the number of devices connected to the Internet create a multitude of social data. These social data pose great challenges in terms of their analysis by companies. With the development of new technologies and new digital communication processes, it is imperative to enhance the current understanding of the possibilities offered by social data for brand communication analysis in the financial sector.

In this context, the present study has compared two research methodologies, the PEI method and the sentiment analysis. Sometimes, social data generated by users are difficult to obtain. With Sentiment Analysis technique, data collection uses mechanisms based on machine learning to download and analyze Tweets [37], [38], which allows a quicker data understanding and in a clearer way than conducting traditional interviews, in which the analysis process is slower and there are qualitative biases based of the researcher behavior.

According to our results, the two compared methodologies yield highly comparable results. However, the methodology based on Big Data analytics that relies on machine learning methods allows us to analyze the data faster than when the traditional method is used.

Our results have important theoretical implications. First, comments made on social networks such as Twitter can help better understand the financial industry sector. Second, the sentiment-analysis-based methodology used in the present study can be meaningfully applied to the analysis of large amounts of data collected from different sources. Thirdly, in the present study, we used the innovative method and identified patterns and indicators that were previously unexplored [4], [38].

In the future, the methodological approach proposed in the present study can be used in further research on data analysis in the financial sector [3], [4].

Finally, our results provide meaningful practical insights for professionals. Based on the knowledge about the image that financial companies have among social platform users, managers of those companies can take advantage of this research study results in order to improve the reputational image of their companies.

REFERENCES

[1] C. Chen, *Technology and Learning: Theory and Practice*. Taipei, Taiwan: Psychology Press, 2004.

[2] A. E. Grant and J. H. Meadows, *Communication Technology Update*. Amsterdam, The Netherlands: Elsevier, 2006.

[3] C. Antoun, "Who are the Internet users, mobile Internet users, and mobile mostly Internet users? Demographic differences across Internet-use subgroups in the U.S," in *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*, D. Toninelli, R. Pinter, and P. de Pedraza, Eds. London, U.K.: Ubiquity Press, pp. 99–117. doi: 10.5334/bar.g.

[4] J. R. Saura, P. Palos-Sanchez, and A. Grilo, "Detecting indicators for startup business success: Sentiment analysis using text data mining," *Sustainability*, vol. 11, no. 3, p. 917, 2019.

[5] N. Golchha, "Big data-the information revolution," *Int. J. Adv. Res.*, vol. 1, no. 12, pp. 791–794, 2015.

[6] B.-G. Jang and H.-S. Lee, "Retirement with risk aversion change and borrowing constraints," *Finance Res. Lett.*, vol. 16, pp. 112–124, Feb. 2016. doi: 10.1016/j.frl.2015.10.003.

[7] J. L. Orihuela and A. Cambronerom, *La Revolución de Los Blogs: Cuando Las Bitacoras se Convirtieron en el Medio de Comunicacion de la Gente*. Madrid, Spain: La esfera de los libros, 2006.

[8] M. Blanco-Callejo, "Banco Santander, en la 'pole position' de las marcas financieras internacionales," *Globalización, Competitividad Gobernabilidad Georgetown/Universia*, vol. 4, no. 2, pp. 7–11, 2010.

[9] J. Amis, N. Pant, and T. Slack, "Achieving a sustainable competitive advantage: A resource-based view of sport sponsorship," *J. Sport Manage.*, vol. 11, no. 1, pp. 80–96, 1997.

[10] R. M. Grant, "The resource-based theory of competitive advantage: Implications for strategy formulation," *California Manage. Rev.*, vol. 33, no. 3, pp. 114–135, 1991.

[11] J. Cerviño, "La globalización de las marcas españolas: Estrategia internacional e imagen de marca del santander," *Universia Bus. Rev.*, vol. 17, pp. 60–73, 2008.

[12] I. Küster, N. Vila, J. Aldás, and C. Rodriguez, "Efecto del patrocinio de la Copa América en las perspectivas de Luis Vuitton: Una perspectiva internacional," *Universia Bus. Rev.*, vol. 22, pp. 40–55, 2009.

[13] E. Ortega, *Determinación de la Imagen de Marca a Través del Método EPI, Grandes Temas de Marketing*. Madrid, Spain: ESIC Editorial, 1989.

[14] B. H. Martinez *et al.*, *Los Mercados del Mañana: Bases Para SU Análisis Hoy*. Cuadernos de Turismo, ESIC Editorial, 2011

[15] E. Ortega, B. Martínez, B. R. Herráez, and S. Christofle, "El método EPI aplicado a la determinación de la imagen de las capitales europeas," *Cuadernos Turismo*, vol. 29, pp. 161–182. [Online]. Available: http://revistas.um.es/turismo/article/view/153831

[16] E. M. Salinas, "La imagen corporativa de las entidades financieras en Aragón," *Documento de Trabajo-Fundación Economía Aragonesa*, p. 68, 2008.

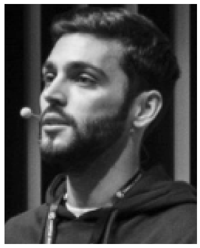
[17] C. Acevedo and J. Nohara, *Monografia no Curso de Administracão: Guia Completo de Conteúdo e Forma*, São Paulo, Brazil: Atlas, 2010.

[18] R. S. Sérgio, T. P. Christopoulos, and E. P. V. Prado, "Comportamento dos bancos no *twitter* e seus efeitos sobre a imagem da marca," *Revista Gestao*, vol. 24, no. 1, pp. 2–12, 2017. doi: 10.1016/j.rege.2016.07.001.

[19] F. Mattar, *Pesquisa de marketing: Metodologia e Planejamento*, São Paulo, Brazil: Atlas, 1996.

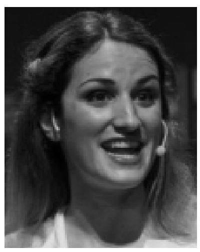
[20] B. Shahzad *et al.*, "Quantification of productivity of the brands on social media with respect to their responsiveness," *IEEE Access*, vol. 7, pp. 9531–9539, 2019. doi: 10.1109/ACCESS.2019.2891081.

[21] S. F. S. Alwi and R. V. Da Silva, "Online and offline corporate brand images: Do they differ?" *Corporate Reputation Rev.*, vol. 10, no. 4, pp. 217–244, 2007. doi: 10.1057/palgrave.crr.1550056.
[22] S. Chakraborty and I. Chung, "Challenges and opportunities of omnichannel retailing," *Eur. J. Risk Regulation*, vol. 5, no. 3, pp. 386–388, 2014. doi: 10.1017/s1867299x00003937.
[23] E. M. Payne, J. W. Peltier, and V. A. Barger, "Omni-channel marketing, integrated marketing communications and consumer engagement: A research agenda," *J. Res. Interact. Marketing*, vol. 11, no. 2, pp. 185–197, 2017. doi: 10.1108/JRIM-08-2016-0091.
[24] J. R. Saura, P. R. Palos-Sanchez, and M. A. R. Martin, "Attitudes to environmental factors in the tourism sector expressed in online comments: An exploratory study," *Int. J. Environ. Res. Public Health*, vol. 15, no. 3, pp. 553–566, 2018. doi: 10.3390/ijerph15030553.
[25] M. Palomino *et al.*, "The online dissemination of nature–health concepts: Lessons from sentiment analysis of social media relating to 'nature-deficit disorder,'" *Int. J. Environ. Res. Public Health*, vol. 13, no. 1, p. 142, 2016. doi: 10.3390/ijerph13010142.
[26] C. C. Ekenga, C.-A. McElwain, and N. Sprague, "Examining public perceptions about lead in school drinking water: A mixed-methods analysis of Twitter response to an environmental health hazard," *Int. J. Environ. Res. Public Health*, vol. 15, no. 1, p. 162, 2018. doi: 10.3390/ijerph15010162.
[27] A. Reyes-Menendez, J. R. Saura, P. R. Palos–Sanchez, and J. Alvarez-Garcia, "Understanding user behavioral intention to adopt a search engine that promotes sustainable water management," *Symmetry*, vol. 10, no. 11, pp. 584–596, 2018.
[28] J. R. Saura, A. Reyes-Menendez, and C. Alvarez-Alonso, "Do online comments affect environmental management? Identifying factors related to environmental management and sustainability of hotels," *Sustainability*, vol. 10, no. 9, p. 3016, 2018. doi: 10.3390/su10093016.
[29] D. Bennett, D. P.-B. Yábar, and J. R. Saura, "University incubators may be socially valuable, but how effective are they? A case study on business incubators at universities," in *Entrepreneurial Universities* (Innovation, Technology, and Knowledge Management), M. Peris-Ortiz, J. Gómez, J. Merigó-Lindahl, and C. Rueda-Armengot, Eds. Cham, Switzerland: Springer, 2017. doi: 10.1007/978-3-319-47949-1_11.
[30] P. R. Palos-Sanchez, J. R. Saura, and F. Debasa, "The influence of social networks on the development of recruitment actions that favor user interface design and conversions in mobile applications powered by linked data," *Mobile Inf. Syst.*, vol. 2018, Feb. 2018, Art. no. 5047017. doi: 10.1155/2018/5047017.
[31] P. R. Palos-Sanchez, J. R. Saura, A. Reyes-Menendez, and I. V. Esquivel, "Users acceptance of location-based marketing apps in tourism sector: An exploratory analysis," *J. Spatial Org. Dyn.*, vol. 6, no. 3, pp. 258–270, 2018.
[32] J. R. Saura, A. Reyes-Menendez, and P. Palos-Sanchez, "Un Análisis de Sentimiento en twitter con machine learning: Identificando el sentimiento sobre las ofertas de #BlackFriday. A sentiment analysis in Twitter with machine learning: Capturing sentiment from #BlackFriday offers," *Revista Espacios*, vol. 39, no. 16, p. 75, 2018.
[33] *MonkeyLearn API Reference*. Accessed: Jun. 15, 2017. [Online]. Available: https://monkeylearn.com/docs/article/api-reference/
[34] J. R. Saura, P. R. P. Sánchez, and A. Reyes-Menendez, " Marketing a través de aplicaciones móviles de turismo (m-tourism): Un estudio exploratorio," *Int. J. World Tourism*, vol. 4, no. 8, pp. 45–56. 2007.
[35] W.-S. Kwon and S. J. Lennon, "What induces online loyalty? Online versus offline brand images," *J. Bus. Res.*, vol. 62, no. 5, pp. 557–564, 2009. doi: 10.1016/j.jbusres.2008.06.015.
[36] A. Immonen, P. Pääkkönen, and E. Ovaska, "Evaluating the quality of social media data in big data architecture," *IEEE Access*, vol. 3, pp. 2028–2043, 2015.
[37] R. Vatrapu, R. R. Mukkamala, A. Hussain, and B. Flesch, "Social set analysis: A set theoretical approach to Big Data analytics," *IEEE Access*, vol. 4, pp. 2542–2571, 2016.
[38] I. Perikos and I. Hatzilygeroudis, "A framework for analyzing big social data and modelling emotions in social media," in *Proc. IEEE 4th Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Mar. 2018, pp. 80–84.



JOSE RAMON SAURA received the International Ph.D. degree in business economics from Rey Juan Carlos University, Spain, Madrid, London South Bank University (LSBU), and Harvard University (HBS), the Senior Management master's degree from Rey Juan Carlos University, and the Online Reputation and Digital Strategy master's degree from IEBS University, Barcelona, Spain. He has more than six years of experience in providing consultancy services in digital marketing companies and start-ups. He has made research stays at the University of Algarve, the Universidade NOVA de Lisboa, and Real Colegio Complusente (RCC), HBS. He is currently a Professor with the Department of Business Economics, lecturing at the Faculty of Social Science and Law of Madrid. He is specialized in sustainable business, digital marketing, social networks, and digital tourism. He has published more than 11 articles indexed in the Web of Sciences (JCR) and the Scopus rankings, among others. He has published chapters in Springer Editorial, IGI Books, and NOVA Publishers. His research interests include digital marketing, search engines, technology adoption, mobile applications, social media, and sentiment analysis. He was a Reviewer for internationals congresses, such as Tourism and Management Studies International Conference, Algarve, Portugal, the III International Conference on Economic and Business Management (FEBM), Hohhot, China, in 2018, and the European Academy of Management and Business Economics, Madrid, Spain, and Gandia, Spain, in 2017 and 2018, respectively. He has been a Reviewer for journals indexed in the Web of Sciences Ranking (JCR) in editorials, such as Elsevier, Wiley, Springer, Inderscience, and MDPI.

BEATRIZ RODRÍGUEZ HERRÁEZ received the Ph.D. degree in business economics. She is currently a Marketing Professor with Faculty of Social Sciences, Rey Juan Carlos University, Madrid. Her research interests include the tourism sector and the development of business plans for brand management. She also develops a line of research regarding the reputation of financial companies and the brand value of their plans and tactics. She is also a Reviewer of indexed journals within the impact ranking as well as an Editor of various special issues and reviewer in the International Congresses of Relevance and Prestige.



ANA REYES-MENENDEZ received the Senior Management master's degree and the Ph.D. degree in business economics from Rey Juan Carlos University, Madrid, Spain. She has made Postdoctoral research stays at Universidade Portucalense, Portugal, and the Harvard Business School, USA. She is currently a Professor with the Department of Business Economics, lecturing at the Faculty of Social Science and Law of Madrid. She is specialized in sustainable business, digital marketing, social networks, and tourism. She has published articles indexed in the Web of Sciences (JCR) and the Scopus rankings, among others. She has published a chapter in INJUVE and AAP Publishers. Her research interests include digital marketing, neuromarketing, search engines, technology adoption, mobile applications, and sustainable business. She is a Reviewer for journals indexed in JCR in editorials, such as Emerald, Inderscience, and MDPI. She has also been a Reviewer for internationals congresses, such as the Tourism and Management Studies International Conference, Algarve, Portugal, and the III International Conference on Economic and Business Management (FEBM), Hohhot, China, in 2018.