RESEARCH ARTICLE

European Journal of Information Technologies and Computer Science
www.ej-compute.org

Big Data Analysis in Commercial Social Networks: Analysis of Twitter Reviews for Marketing Decision Making

Imane Satauri, Boutaina Satouri, and Omar El Beqqali

ABSTRACT

Content generated by users on commercial social networks about products and brands generates large volumes of data that can be transformed into relevant and useful recommendations for marketing decisions. Every day, consumers post their opinions online on social networks about products they have purchased and used, and companies are increasingly interested in tracking this information in real time for better decision making. The main problem is to extract key information from consumers' textual comments and use it automatically to measure the quality of products or brands. In this work, we propose a hybrid approach to automatically analyze these reviews, assigning a quantitative score to negative and positive user content.

The analysis of online consumer sentiment has increased significantly in recent years, being crucial to determine the success of businesses in a wide range of sectors, tourism, hospitality and e-commerce. In the same context, this work proposes a framework for analyzing the sentiment of reviews posted on the Twitter network towards products and brands. The first step is the construction of a dataset by collecting a set of reviews posted online on Twitter, processing and cleaning the textual data for better accuracy, and then developing a hybrid approach for product evaluation and polarities creation using lexicon-based methods and machine learning-based analysis techniques.

Keywords: Bag-of-words representation, decision making, machine learning, online reviews, sentiment analysis.

Published Online: May 11, 2023

ISSN: 2736-5492

DOI :10.24018/ejcompute.2023.3.2.94

I. Satauri*
Laboratory of Computer Science, Signals, Automation and Cognitivism, FSDM, Morocco.
National School of Commerce and Management, USMBA, FEZ, Morocco.
(e-mail: imane.satauri@usmba.ac.ma)

B. Satouri
Laboratory of Computer Science, Signals, Automation and Cognitivism, FSDM, Morocco.
Sidi Mohamed Ben Abdelah University, FEZ, Morocco.
(e-mail: boutaina.satouri@gmail.com)

O. El Beqqali
Laboratory of Computer Science, Signals, Automation and Cognitivism, FSDM, Faculty of Sciences DM, USMBA, FEZ, Morocco.
(e-mail: omar.elbeqqali@usmba.ac.ma)

Corresponding Author

I. Introduction

Big Data is one of the most important challenges for academics and practitioners. Every day, very large amounts of data from different heterogeneous sources are generated at high speed, which constitutes great difficulties for companies and executives to transform this data into useful information to effectively manage the relationship with their customers. In this sense, massive data analysis has rapidly become a popular practice adopted by many organizations to obtain valuable information [1]. As a result, there are increasing demands for applications that can provide fast and real-time interpretations and predictions of the future to businesses [2], [3].

In marketing, the importance of big data highlights the need to understand how customer data analysis could influence, detect and satisfy their real demands [4].

The study and development of machine learning and big data analytics algorithms for the marketing field is still in its infancy, which makes it essential to increase the efforts for big data to be recognized as a key tool in the marketing field [5].

Among the data streams generated on social networks, we mention user-generated content (UGC) that constitutes a valuable source of information such as blogs, social networks and online consumer reviews), through which Internet users generate different streams of information related to the brand, this information has a decisive commercial value for advertising campaigns [6], for customer relationship management [7] and for brand communication [8].

Among the most used types of UGC, we mention the electronic word-of-mouth (eWOM) expressed by users, it is a very useful technique and effectively influences the purchase decision of customers [9]. As stated by [10], purchase decision making is affected not only by opinions formed by experts, but also by ordinary people. Consumers typically look for reviews and comments written by others when they wish to purchase a product online. Thus, mining and analyzing UGC data such as reviews and sentiments could be useful for companies in the sense that they can influence brand image and positioning.

As noted by [11], companies are increasingly collecting customer feedback data as they are crucial and strongly influence purchase actions. However, regarding the nature of UGC, current work assigns numerical indexes such as rating scores [12], other research works use text mining to examine textual content of UGC and assigning categories to the reviews (e.g., [13]; [14]).

Several systems for detecting and analyzing online reviews have been developed with the objective of distinguishing and identifying fake reviews, but it is still very difficult to be sure whether a review is fake or not. Some systems analyze both the review and the user who generated it based on several criteria (number of reviews, reasons for purchase, incompatible dates) and many other indicators for detecting the type of review.

Some companies hire professionals or communication agencies to publish positive reviews of their products to promote their product or attack their competitors' products. Hired spammers are paid according to the published reviews [15]. For example, they use robots to make mass publications, or publish the same or similar content several times.

In this paper, we propose the development of a framework that will enable marketers to interpret qualitative UGC in an automatic way, and transform it into quantitative representations, the design of a review analysis system that combines lexical sentiment analysis tools and machine learning algorithms to help marketers and consumers in their decision-making process. follow.

II. Related Work

A. Detection of Fake Reviews

The problem of detecting fake reviews has been the subject of several research works since 2007, for example the authors [16] studied the case of Amazon and concluded that manual detection of fake reviews can be difficult especially when dealing with large data sources, as fake reviewers may write their reviews carefully in order to make them more reliable for Internet users. They proposed the use of duplicates as spam to develop a model that detects fake reviews [16]. Detecting fake reviews is a specific application of the general problem of fraud detection, for which both verbal and non-verbal metrics can be used [17]. Several research works on the analysis of fake online reviews have exploited textual and behavioral features, other approaches have considered social and temporal features.

Reference [18] studied some psycholinguistic indicators based on LIWC [19] combined with standard words and parts of speech (POS) based on N-gram models. Reference [20] extended this work by also adding style and POS based indicators such as syntax depth and POS sequence patterns. However, fake review detection based on textual indicators alone remains less reliable and difficult. Other authors propose additional textual indicators such as semantic similarity and emotion [21], a wide variety of lexical and syntactic features [22], and more detailed features such as comprehensibility, level of detail, writing style, and other cognition indicators [23].

The behavioral indicators refer to non-verbal characteristics, such as the number of reviews or the time and device where the review was posted. They have been used to improve the reliability of the classification model and the results obtained are encouraging.

Reference [24] introduced behavioral features on Amazon reviews, distinguishing indicators such as (number of feedbacks, review position, rating...), others related to the product such as (price, sales ranking...) and many other indicators related to the author such as (average rating, ratio of the number of reviews written by the review author and that are positive...).

Most of the research done for the problem of detecting fake online reviews has used supervised learning algorithms such as (Support Vector Machine [25], Naive Bayes [26], decision trees [27], random forests [28] and logistic regression [29]), other approaches have been developed, for example in [30], the authors propose a new prediction model based on semi-supervised learning and a set of textual and behavioral indicators

B. Sentiment Analysis

The Sentiment analysis is a recent field of study that uses automatic mechanisms for natural language processing. It concerns the identification of mood, or opinion of subjective elements in a text [31]. With the evolution of the Internet and its applications, textual data is increasingly increasing from many sources. In marketing, for example, consumers post content, provide their reviews online, and generate large volumes of data from social networks, product review websites, blogs, and the Internet. These information flows are generated through platforms where users can freely communicate, exchange ideas, and give their opinions about brands, tourist destinations, restaurants ... [32].

Sentiment analysis for evaluating product quality corresponds to the process of analyzing reviews to identify consumer feelings towards a product [33]. The problems that need to be solved in NLP use sentiment analysis processes and techniques, however the work of [34] shows different types of problems encountered in sentiment analysis such as: 1) sentences that have reverse meaning, 2) interrogative sentences often generate a neutral sentiment, 3) some sentences include information about the sentiment, while they do not use words that indicate sentiment.

The work of [35] has shown the usefulness of Text Mining approaches in sentiment analysis: clustering; web mining; information retrieval; knowledge extraction…

Sentiment analysis mainly focuses on the classification of polarity (positive, negative, neutral) in text data to classify online product reviews. Such an approach is used for exploring consumers' opinions about products. According to [34], this approach can be used to determine the parameters or parts that need to be improved and modified for a product by summarizing the conversations and comments related to it (see which product features generate positive or negative reviews by consumers).

According to the literature review, several application areas use sentiment analysis: commerce for an efficient choice of products based on customer feedback, recommendation systems to analyze user reviews for books, marketing, or also political data analysis [32].

According to [36], machine learning techniques have been used, however fuzzy classifiers are not widely used for this type of problem, due to the nature of the language that makes the classification task difficult for this type of algorithm.

For all these methods, the use of a sentiment lexicon is mandatory, for example SentiWordNet is a WordNet-based parsing lexicon, it assigns to each group of synonyms from WordNet, three sentiment scores: positivity, negativity, and objectivity [37].

We also note word lists related to the affective lexicon such as “SenticNet”, “Afinn” and “WordNet-Affect”.For example, “Afinn” is a powerful word list manually evaluated with a rating score between -5 and +5, giving emotional indices for 2476 English words [38].

III. Proposed Approach

In this section, we present our approach based on the implementation of a Framework for the analysis of online reviews of brands and products, using sentiment analysis tools. Thus, our model can inform and improve decision making for both companies and consumers.

As shown in Fig. 1, the proposed model consists of several steps namely:

(1) data preprocessing, (2) sentiment analysis process, (3) negative and positive information extraction (4) dashboards.

This model takes as input data sources of different types such as online reviews of products on social networks (through web scraping techniques), information about each product such as: price; brand; and category, and many other information. This data is then analyzed to extract relevant information through consumer feedback on product quality, expectations, etc. Then, a hybrid model is defined and computed to form a classifier that detects negative/positive consumer feedback and comments. Thus, new information is extracted to help companies and consumers make the right decisions about products (e.g. purchase decision, improve product quality...).

User reviews are usually in the form of scores and unstructured text comments. The score is a rating that the user chooses from a range of values, for example, from 1 to 5 stars, where a value of 1 means a very poor rating and a value of 5 means a very good return. However, some discrepancies may arise when, for example, the product receives an overall rating of 4 or 5 stars, but the user indicates criticism of specific product features. In this case, the proposed model will adjust the rating accordingly, even if the text review contains both positive and negative index about different aspects of the product.

The proposed system aims to use different sentiment analysis tools already used in previous work. We rely on an architecture that is modular enough to allow the incorporation of various sentiment analysis tools. These include a set of NLP techniques - in particular lexical, syntactic, and semantic analysis as well as machine learning methods.

Some sentiment analysis tools allow you to obtain a quantitative ranking from the textual comments.

The Sentiment score is one of the key indices to detect the best items of a particular product category or brand.

In our approach, fake reviews will not be taken into account. We rely on NLP techniques to recognize similar reviews, e.g. two similar reviews from different users have a high probability of being fake. For each review, we calculate the similarity based on a comparison with all other reviews. A review is classified as a fake review if its similarity is higher than a certain threshold, so we will have a clear view of the product ranking, offering an intelligent market management in real time.

The steps visualized in Fig. 1 are described in detail in the following:

1) Pre-processing of reviews by breaking down each user's comments into words, then assigning lexical, semantic and syntactic labels to each word. These labels further improve the evaluation results.

2) Sentiment analysis, which attempts to assign a label to product reviews, depending on the labels used, that can be either "positive" or "negative".

For natural language processing problems, several libraries and programming languages are available to facilitate the analysis. Among them, we have chosen the R language which is widely used in the field of data analysis. We also used some libraries for text analysis, like TextBlob from python. The idea will be to be able to do sentiment analysis of a tweet based on the words used, using the TextBlob library as a first step.

The structure of a tweet is less organized. Some use emojis, the use of punctuation is much more present, spelling mistakes are very common. All this complicates the preliminary cleaning work, especially when it comes to sentiment analysis. To get around these problems, we used TextBlob. This module has an option to measure the sentiment of a given text. In output the function gives us a polarity coefficient and a sensitivity coefficient. The more the polarity is close to -1 the more the tweet is negative, on the contrary a polarity close to 1 means that the tweet is rather positive. The objective is to be able to follow in real time the satisfaction of the customers, for example if a customer sends an email and the system detects that the person is annoyed, you know that you risk losing a customer.

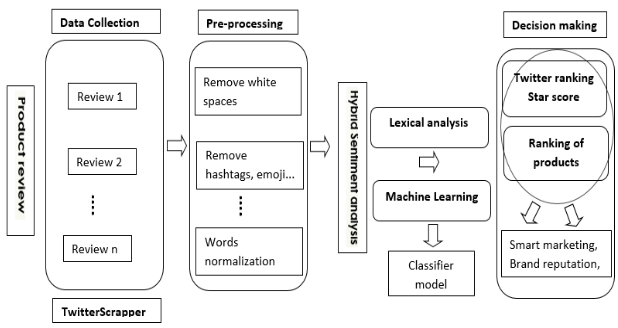

Fig. 1. The proposed architecture for the analysis of online reviews on Twitter.

The dataset used is quite large, with several thousand values. We thought of grouping the tweets in packages according to the chronological order. We will then keep for each group only its average polarity.

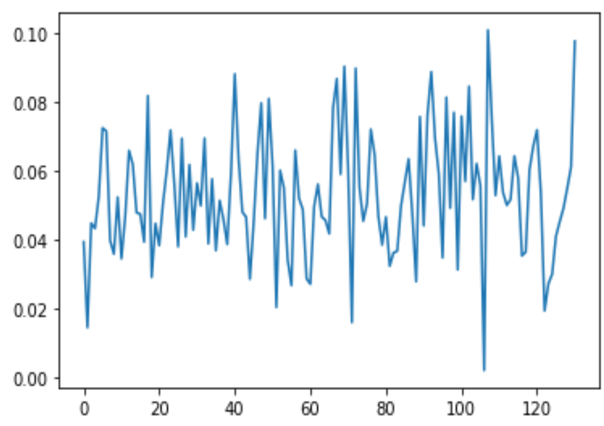

Fig. 2. Evolution of the sentiments of the tweets by packets.

After completing step 2 (sentiment analysis), we will obtain a score for each product. Then, 3) we will perform a series of exploratory analyses that can improve the decision making for marketing, manufacturing/production, ... and also the purchase decision for consumers. The exploratory analysis process will attempt to answer several questions through the information hidden in the big data contained in the reviews published on products. Our framework is based on the score assigned by the user, the score calculated by the sentiment analysis method and also the selling price of the product.

Among these questions we mention for example: What is the best product in each category according to price or consumer satisfaction? What is the product with the highest satisfaction of a specific brand? For a customer looking for the best product at the best price of a specific brand, what is the best product in terms of price and Customer satisfaction, not only according to the number of stars, but also according to customer reviews?

In the last step of our approach (4), we develop multidimensional views that will help companies and consumers in the decision making. With this module, managers can see products that need corrective actions. For example, the sales activity of a product may decrease due to some negative reviews and conversely, a product may have positive feedback but its price is too high to generate new sales.

IV. Experimentation and Results

A. Data Collection

For collecting data, we used the Python module Twitterscrapper to extract a large number of tweets by specifying date and language criteria and by limiting ourselves to tweets that contain certain keywords. The resulting dataset contains reviews and opinions posted on a set of products during the period from January 1 to June 1, 2022. Each review includes a rating from 1 to 5 and a positive/negative description assigned by the User. We also find for each product, a set of data containing information specific to the product, namely the description of the product, information on the category, price, brand ... etc..

Subsequently, we built a data cleaning pipeline with the Python module « re » as follows:

```
import re

def nlp_pipeline(text):

    text = text.lower()
    text = text.replace('\n', ' ').replace('\r', '')
    text = ' '.join(text.split())
    text = re.sub(r"[A-Za-z\.]*[0-9]+[A-Za-z%°\.]*", "", text)
    text = re.sub(r"(\s\-\s|-$)", "", text)
    text = re.sub(r"[,\!\?\%\(\)\/\"]", "", text)
    text = re.sub(r"\&\S*\s", "", text)
    text = re.sub(r"\&", "", text)
    text = re.sub(r"\+", "", text)
    text = re.sub(r"\#", "", text)
    text = re.sub(r"\$", "", text)
    text = re.sub(r"\£", "", text)
    text = re.sub(r"\%", "", text)
    text = re.sub(r"\:", "", text)
    text = re.sub(r"\@", "", text)
    text = re.sub(r"\-", "", text)

    return text
```

Fig. 3. Pipeline for cleaning up online reviews (tweets).

This pipeline allowed us to get tweets that were pretty much clean. This allows TextBlob to analyze the tweets more efficiently.

A	B	C	D	E	F	G	H	I	J	K
	id	conversation_id	created_at	date	timezone	place	tweet	hashtags	cashtags	user_id
0	122683821670529	122680049225851	1581335999000	2020-02-10 11:59:5	UTC		Jcrois il repousse le	[]	[]	714993033884909
1	122683821034673	122654399526806	1581335997000	2020-02-10 11:59:57	UTC		Et on nous dit de lim	[]	[]	121035123233179
2	1226838134744326	1226838134744326	1581335979000	2020-02-10 11:59:3	UTC		ALERTE - Le Roy	['#coronavirus', '#cor	[]	121431561903147
3	122683813414454	122683813414454	1581335979000	2020-02-10 11:59:3	UTC		Ptdrr ils nettoyent sa	['#contaminesmontj	[]	362050772
4	122683809474751	122683809474751	1581335969000	2020-02-10 11:59:2	UTC		Coronavirus : le bilar	[]	[]	378817170
5	1226838073595613	122683689040763	1581335964000	2020-02-10 11:59:24	UTC		tout à fait d'accord a le pire c'est que cert	[]	[]	2162947538
6	122683804199995	122683804199995	1581335957000	2020-02-10 11:59:17	UTC		On est lundi, la réuni on sent l'urgence . #CORONAVIRUSENI #CORONAVIRUS . https://www.lefigaro.	['#coronavirusenfran	[]	57597804
7	122683804185314	122683804185314	1581335957000	2020-02-10 11:59:17	UTC		Football : La mauvai	[]	[]	85862615
8	122683797889663	122683797889663	1581335942000	2020-02-10 11:59:0	UTC		Une équipe de l'OM	['#senegal', '#kebetu'	[]	3918185062
9	[illegible]	[illegible]	[illegible]	[illegible]	UTC		Droit aux indemnités	[]	[]	[illegible]

Fig. 4. Extract from the tweets dataset.

B. Experimentation

- *Annotation creation and sentiment analysis*

Text Blob takes the text as input and returns the polarity and subjectivity as output. The polarity determines the sentiment of the text whose values are in the interval [-1,1] where -1 denotes a very negative sentiment and 1 denotes a very positive sentiment.

We show in the following, the calculation of the polarity of feelings through the example below:

```
from textblob import TextBlob
t_1 = "smells amazing, a perfect purchase. Nothing really negative for the product, except for the temperature , and it's just because of my fine hair… !."
t_2 = "Must buy , super amazing ☺ "
t_3 = "Quite satisfactory "
pl_1 = TextBlob(t_1).sentiment.polarity
pl_2 = TextBlob(t_2).sentiment.polarity
pl_3 = TextBlob(t_3).sentiment.polarity
s_1 = TextBlob(t_1).sentiment.subjectivity
s_2 = TextBlob(t_2).sentiment.subjectivity
print("Polarity of Text 1 is", pl_1)
s_3 = TextBlob(t_3).sentiment.subjectivity
print("Polarity of Text 2 is", pl_2)
print("Subjectivity of Text 1 is", s_1)
print("Subjectivity of Text 2 is", s_2)
```

Fig. 5. Extract from the polarities calculation program.

The result is as follows:

```
Polarity of Text 1 is 0.9
Polarity of Text 2 is 1.0
Polarity of Text 3 is 1.0
Subjectivity of Text 1 is 1.0
Subjectivity of Text 2 is 1.0
Subjectivity of Text 3 is 1.0
```

Fig. 6. Result of the polarities calculation.

The results are satisfactory but sometimes some specific details are not well examined while the star scores given by the users are high. For example, on Text 1: “Nothing really negative for the product, except for the temperature, and it’s just because of my fine hair…!” (The polarity obtained by our model is 0.9 and the score assigned by the user is 4). The consumer can express different opinions in the text review, however some impressions are not translated into the numerical score.

In order to improve the accuracy of our analysis system and also to adjust the correlation between sentiment values and star scores assigned by consumers, we propose a hybrid model that combines the lexicon-based approach presented above that allowed us to make an initial annotation of the tweets and a supervised learning approach with machine learning algorithms to have as output a classifier of the polarity of tweets. This combination allowed us to significantly improve the accuracy.

- *Machine learning for the prediction of polarities*

Our model was able to identify 1765 features. These descriptors play an important role in the classification of sentiments. For this purpose, we tested several combinations of extraction and weighting schemes to ensure the best quality of the developed models. After studying several representations, we chose the bag of words model.

In practice, it is always recommended to compare the performance of at least a few learning algorithms in order to select the best model for the problem at hand. These may differ by the number of descriptors or samples, the amount of noise in a dataset and whether the classes are linearly separable or not.

In this context, we have applied the most widely used classifiers in the literature to sentiment analysis. In particular, many classifiers have been applied on our Dataset. Regarding the implementation, we chose the python library Sklearn for the execution of the used classifiers.

After the learning phase, we move to the testing phase to evaluate our classifier. For performance validation, we used the 70% to 30% rule to validate our model, such that 70% of the data is reserved for the learning phase and 30% for the testing phase. The combination of extraction and weighting schemes allowed us to test different configurations. The following table summarizes the results of the experiments conducted:

TABLE I: RESULTS OF THE MACHINE LEARNING ALGORITHMS FOR THE CLASSIFICATION OF TWITTER REVIEWS.

Classifier	Number of Descriptors Used	Representation With «Bag of Words»
Naïve Bayes	1765	0,909
Decision Tree	1765	0,945
Random Forest	1765	0,948

For each algorithm, we present the rate of good ranking (Accuracy), obtained with the test sample. In general, these results show that the best performance was obtained with the [KNN -bag of words] combination. The following table shows the results of the classifiers for the bag of words representation model:

TABLE II: CLASSIFIER RESULTS FOR THE "BAG OF WORDS" REPRESENTATION MODEL.

Classifier	Accuracy (¨%)	Recall(%)
Naïve Bayes	93,2	98,2
Decision Tree	98,86	95,92
Random Forest	98,72	93,87
Logistic Regression	94	95,66
KNN	99,10	95,6

- *Exploring the results*

The main objective of our approach is to assist clients and organizations in the decision making related to the evaluation of feedbacks and comments on products and brands. In this section, we detail the exploratory analysis stage of the results obtained through the sentiment analysis process to provide real-time information and recommendations to guide purchasing actions and help companies to properly conduct marketing strategies.

For example, if we want to answer the following query:" What is the most requested product based on price and/or customer satisfaction?", Our approach can easily answer this query and give the best products in each category according to the normalization of the star scores with the sentiment values calculated by our system. We use different parameters: (1) the number of stars, (2) selling price of the product, and (3) the sentiment polarity to analyze the correlation between the price and the quality of the products, we assigned a weight to each variable to get the average score and then specify the good product. This score is obtained from the weighted average of the normalized price(weight=0.5), the star score and the sentiment value(weight=0.4).

TABLE III: CLASSIFICATION OF PRODUCTS ACCORDING TO THE PROPOSED MODEL

Product	Category	Price	Star score	Sentiment value	Ranking
Fragnance 1 (Lavender)	Perfume	P1	S1	1.79	0.89
Fragnance 2 (Rose)	Perfume	P2	S2	1.76	0.78
Fragnance 1 (Lemon)	Perfume	P3	S3	0.45	0.56

In Table III, we have calculated the average score for three products in the category "Fragrance". We notice that the good product, Fragrance1(Lavender), is not expensive, but it often receives favorable reviews by users based on the polarity of sentiment and the number of stars.

V. CONCLUSION

In this work, we have proposed a modular approach based on sentiment analysis to help companies and consumers in the decision making process. It provides relevant information extracted from consumer reviews and processes it using NLP technologies to evaluate product/brand quality and analyze user feedback. We have deeply studied previous works on big data issues in marketing, sentiment analysis and also online fake review analysis, and the results are summarized in the "state of the art" section.

To implement our approach, we used an extract of reviews published on the Twitter network to evaluate products according to the feedback and comments posted by

European Journal of Information Technologies and Computer Science
www.ej-compute.org

consumers. We noticed in some cases, differences between the sentiment values obtained by our model and the number of stars assigned by the users, which shows that textual feedbacks may contain hidden and relevant information that is not clearly expressed in the star scores.

In this way, our approach will allow decision makers to make comparative analyses to answer important and useful questions for the study of customer behavior, such as " What is the most requested product based on price and/or customer satisfaction for each category or for a specific category?" or "What is the most appreciated product for a given brand?"

Despite all the existing sentiment analysis tools and despite all the efforts of researchers in this field, the challenges related to NLP are still difficult to solve. The construction of reliable and flexible semantics is not yet fully mastered. We plan to study in future works, how fake reviews influence the accuracy of product evaluation.

On the other hand, the approach proposed in this paper has some limitations concerning the generalization of the results. For example, the "experimentation" part has been applied only on the social network Twitter. Although it is a large commercial platform with a great popularity in different countries, but if the study had been conducted on other platforms such as "AliExpress, Amazon ...", other conclusions could have been obtained.

Finally, the creation of an annotated corpus of fake reviews can help the research community to develop tools that integrate the detection of fake reviews.

Conflict of Interest

Authors declare that they do not have any conflict of interest.

References

[1] Sivarajah U, Kamal M, Irani Z, Weerakkody V. Critical analysis of big data challenges and analytical methods. *Journal of Business Research*. 2017; 70: 263-286.

[2] Lytras M, Raghavan V, Damiani E. Cognitive computing and big data analytics research: From metaphors to value space for collective wisdom in human decision making and smart machines. *International Journal on Semantic Web and Information Systems*. 2017; 13(1): 1-10.

[3] Chen H, Chiang R, Storey V. Business intelligence and analytics: From big data to big impact. *Management Information Systems Quarterly*. 2012; 36(4): 1165-1188.

[4] Kumar A, Shankar R, Aljohani N. A big data driven framework for demand driven forecasting with effects of marketing-mix variables. *Industrial Marketing Management*. 2019

[5] Amado A, Cortez P, Rita P, Moro S. Research trends on big data in marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics*. 2018; 24(1): 1-7.

[6] Zhang K, Katona Z. Contextual advertising. *Marketing Science*. 2012; 31(6): 980-994.

[7] Chuang SH. Co-creating social media agility to build strong customer-firm relationships. *Industrial Marketing Management*. 2019.

[8] De Vries L, Gensler S, Leeflang P. Popularity of brand posts on brand Fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing*. 2012; 26(2): 83-91.

[9] Wang Z, Fujita H, Liu S. Towards felicitous decision making: An overview on challenges and trends of big data. *Information Sciences*. 2016; 367: 747-765.

[10] Feldman R. Techniques and applications for sentiment analysis. *Communications of the ACM*. 2013; 56(4): 82-89.

[11] Liu B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies.* 2012; 5(1): 1-167.

[12] Netzer O, Feldman R, Goldenberg J, Fresko M. Mine your own business: Market-structure surveillance through text mining. *Marketing Science*. 2012; 31(3): 521-543.

[13] Homburg C, Ehm L, Artz M. Measuring and managing consumer sentiment in an online community environment. *Journal of Marketing Research.* 2015; 52(5): 629-641.

[14] Liu X, Burns A C, Hou Y. An Investigation of Brand-Related User-Generated Content on Twitter. *Journal of Advertising*. 2017; 46(2): 236-247.

[15] Day M, Wang C, Chen C, Yang S . Exploring review spammers by review similarity: A case of fake review in Taiwan. *Proceedings of the third international conference on electronics and software science (ICESS2017)*, pp.166, 2017.

[16] Nitin J, Bing L. Review spam detection. *In Proceedings of the 16th international conference on World Wide Web*, 2007, pp.1189-1190.

[17] Eileen F, Joan B, Tommaso F. Automatic detection of verbal deception. Synthesis Lectures on Human Language Technologies. 2015; 8(3): 1-119.

[18] Myle O, Yejin C, Jeffrey T. Finding deceptive opinion spam by any stretch of the imagination. *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human LanguageTechnologies*. 2011; 1: 309-319.

[19] Yla R, James W. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*. 2010; 29(1): 24-54.

[20] Arjun M, Vivek V, Bing L, Natalie S. What Yelp fake review filter might be doing?. *ICWSM*, 2013: 418.

[21] Yuejun Li, Xiao F, Shuwu Z. Detecting fake reviews utilizing semantic and emotion model. *In 3rd International Conference on Information Science and Control Engineering (ICISCE)*, 2016, pp. 317-320.

[22] Rupesh K, Dewang, Singh AK. Identification of fake reviews using new set of lexical and syntactic features. *In Proceedings of the Sixth International Conference on Computer and Communication Technology*, 2015, pp. 115-119.

[23] Snehasish B, Alton C, Jung-Jae K. Using supervised learning to classify authentic and fake online reviews. *In Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*, 2015, pp. 88.

[24] Nitin J, Bing L. Opinion spam and analysis. *In Proceedings of the International Conference on Web Search and Data Mining*, 2008, pp. 219-230.

[25] Vladimir V, Steven E, Alex J. Support vector method for function approximation, regression estimation and signal processing. *In Advances in neural information processing systems*, 1997, pp.281-287.

[26] Nir F, Dan G, Moises G. Bayesian network classifiers. *Machine Learning*.1997; 29(2-3): 131-163.

[27] Breiman L, Friedman J, Charles J, Stone, Richard A. Classification and regression trees. CRC press, 1984.

[28] Breiman L. Random forests. *Machine Learning*. 2001; 45(1): 5-32.

[29] David R. The regression analysis of binary sequences. *Journal of the Royal Statistical Society*.1958: 215-242.

[30] Fangtao Li, Minlie H, Yang Yi, Xiaoyan Z. Learning to identify review spam. *In IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, 2011, pp. 2488.

[31] Bhadane C, Dalal H, Doshi H. Sentiment analysis: Measuring opinions. *Procedia Computer Science*. 2015; 45: 808-814.

[32] Rahman K, Khamparia A. Techniques, applications and challenges of opinion mining. *International Journal of Control Theory and Applications*. 2016; 9(41): 455-461.

[33] Haddi E, Liu X, Shi Y. The role of text pre-processing in sentiment analysis. *Procedia Computer Science*. 2013; 17: 26-32.

[34] Jandail R. A proposed novel approach for sentiment analysis and opinion mining. *International Journal of UbiComp*. 2014; 5: 1-10.

[35] Demoulin N, Coussement K. Acceptance of text-mining systems: The signaling role of information quality. *Information & Management*. 2018.

[36] Jefferson C, Liu H, Cocea M. Fuzzy approach for sentiment analysis. *IEEE international conference on fuzzy systems (FUZZ-IEEE)*, 2017, pp.1-6.

[37] Hung C, Lin H. Using objective words in SentiWordNet to improve word-ofmouth sentiment classification. *IEEE Intelligent Systems*. 2013; 28: 47-54.

[38] Nielsen F. A new evaluation of a word list for sentiment analysis in microblogs. *Proceedings of the workshop on 'making sense of Microposts': big things come in small packages 718 in CEUR workshop proceedings*, 2011, pp. 93-98.