Spanish Twitter Data Used as a Source of Information About Consumer Food Choice

Luis G. Moreno-Sandoval, Carolina Sánchez-Barriga, Katherine Espíndola Buitrago, Alexandra Pomares-Quimbaya, Juan Carlos Garcia

Spanish twitter data used as a source of information about consumer food choice

Luis G. Moreno-Sandoval [1,2], Carolina Sánchez-Barriga [1,2], Katherine Espíndola Buitrago [1,2], Alexandra Pomares-Quimbaya [2] and Juan Carlos Garcia [2]

[1] Colombian Center of Excellence and Appropriation on Big Data and Data Analytics (CAOBA), Bogotá, Colombia
[2] Pontificia Universidad Javeriana, Bogotá, Colombia

Abstract. Food related consumer behavior is a topic of major interest to areas such as health and marketing. Social media offers a scenario in which people share information about preferences, interests and motivations about eating habits and food products that have not been exploded as appropriate. In this work we present an algorithm to exploit the potential of Twitter as a data gathering platform to provide insight about behavior of consumers, by linking the food-related content, including emoji's expressed by Twitter users, to their demographic profile (age, gender, socioeconomic level). We further link this data to dietary choices expressed in different moments of their daily life. We found out that including Spanish Twitter data analysis, like the one presented in this work, into marketing researchers tools, could be very useful to advance in customer-centric strategies.

Keywords: Social networks, Consumer Behavior, Twitter, Food Analysis, Social Media.

1   Introduction

1.1   Motivation

Twitter data have become a source of insights to study consumer behavior in different contexts and domains. Regarding food consumer behavior, the need for using Twitter raises in some limitations of common practices due to the strong influence of contextual variables and the subjective charge that consumers' responses can have when looking for socially desirable or over- rationalized answers [1].

However, Twitter offers an option to those limitations, since it becomes a natural consumption setting which provides access to consumer information spontaneously. According to Vidal et al., [1], Twitter is used to present daily information including consumption routines and comments; and, since eating and drinking are some of the most common human activities, tweets can be a data source for food-related consumer behavior insights.

According to Nielsen [2], Twitter is well positioned to study eating situations since it provides researchers the opportunity to retrieve spontaneous data, generated in real-life settings. In this sense, it is possible to collect data in any situation, considering that consumers are increasingly using smartphones to access social media [2].

1.2   The potential of Twitter in marketing research

Evolution of digital trends such as social networks, mobile technologies, cloud computing or Big Data provide a huge source of information about consumer behavior, needs and requirements [3]. Therefore, companies can offer completely new services, participate interactively with customers and provide a completely different work environment, which is the reason why digital technology plays a critical role in consumer research strategies.

The proposal of Uhl et al., [4], argues that using a Customer Centralization strategy on organizations allows to position and to achieve economic success on organizations. From the perspective of the customer's life cycle, the Customer Centralization strategy focuses on four phases: information, supply, purchase and after-sales. In a transversal way, the customer experience goes through every phase. It is defined as a managerial philosophy that involves a complete alignment of the company with existing and potential relationships with its customers; it makes customer the focus of all commercial considerations. The central principle is to increase the value of the company and the customer itself through the systematic management of these relationships.

In order to improve company's customer knowledge and considering Twitter as an invaluable source of information about customer profile, interests and behaviors, this paper proposes an algorithm able to analyze food mention behavior in social networks from different points of view. The algorithm is able to identify the context in which a customer mentions food related words and characterizes the situation in which it was posted. In addition, it considers not only narrative texts, but also hashtags and user mention. The proposed algorithm demonstrates improvement on current approaches employed in food-related studies in social media.

This paper is organized as follows: In the section 2, recent studies that use twitter information to perform analysis in specific areas of knowledge such as food-consumption, tobacco consumption and healthcare are presented. Section 3 describes the proposed algorithm. First, it describes the food extraction processes used for the construction of the knowledge base that supports the algorithm, and then explains the proposed algorithm for food detection. Section 4 presents the main results obtained in the case study and concludes with section 5 presenting our main contributions and future work.

2   Related Works

Recent food-related studies have focused on problems, topics and consumer behavior research in public health. Vidal et al., [1] found that "*people tended to mainly tweet about eating situations when they enjoyed them, due to the foods they consumed, the place in which they were and/or the people they shared the meal with.*"

Abbar et al, [5] found that foods mentioned in daily tweets of users are predictive of the national obesity and diabetes statistics, showing how the calories tweeted are linked to user interests and demographic indicators, and that users sharing more friends are more likely to display a similar interest towards food. This work includes demographic indicators correlated with food-related information. The studies from Abbar et al enriched data using a variety of sources, which allowed considering nutritional value of the foods mentioned in tweets, demographic characteristics of the users who tweet them, their interests, and the social network they belonged to.

In a recent study, Prier et al., [6] used LDA to find topics related to tobacco consumption, such as addiction recovery, other drug use, and anti-smoking campaigns. Finally, Dredze et al., [7] applied a Food Topic Aspect Model on tweets, to find out mentions of various aliments; The results suggest that chronic health behaviors, such as tobacco use, can be identified and measured, however, this does not apply to other short-term health events, such as outbreaks of disease. Also, it is found that the demographics of Twitter users can affect this type of studies, leaving the debate open.

Users of online social networks (OSN) reveal a lot about themselves; however, depending on their privacy concerns, they also choose not to share details that seem sensitive to them, reconfiguring access to their information in the OSN [11]. Many applications on Facebook that are well-known for being able to use them, request a lot of information from the user [12]. On the contrary, the proposal presented in this article is based exclusively on the publicly available information of users of social networks and, in that sense, does not violate any agreement on the use of data applicable in America and Europe. In addition, demographic data derived from OSN users, and employed for the food-consumption analysis, is the result of a previous project that demonstrates and validates the potential of twitter public publications to infer valuable information about its users [13] [14].

3   Consumer food choice identification

In order to explore Twitter data, we used *bag of words* [8] [9] [13] as a method to understand the tweet content related to food consumption. This method uses an initial food knowledge base with 1128 words, generated by an automatic domain constructor [16]. In the first approach of the analysis, we found out that a large portion of the tweets that include food words are not referring to actual food consumption, this is one of the

most important challenges on the algorithm. Most of them refer to popular sayings that include food words like:

*"amigo el ratón del **queso** (friend the mouse of the **cheese**)"*.
*"cuentas claras y el **chocolate** espeso (bills clear and **chocolate** thick)"*.
*"sartén por el **mango** (taking the frying pan by the handle)"*.
*"al **pan, pan** y al **vino, vino** (the **bread** is **bread** and the **wine** is **wine**)"*.

Some tweets had another type of expressions widely used in other contexts, that includes food words; for example, the word ***jam*** in the Colombian political context is associated with corruption issues and is widely used in social networks, for example:

*"Desastroso es un gobierno lleno de **mermelada** y clientelismo (a government full of **jam** and clientelism is disastrous)"*.

Additionally, this tweet understanding also shows that many users refer to specific products associated with popular brands, without using the food word, such as *"Pony Malta (soda)", "Coca – Cola (soda)", "Galletas Oreo (cookies)"*. In the same way, other users use hashtags like *"#almuerzo (#lunch) and #aguacate (#avocado)"* or mentions (or usernames) such as *"@baileysoriginal and @BogotaBeerCo"* to make a reference to products or places of consumption. Finally, we also concluded that users refer to food consumption with emojis as shown on figure 1.
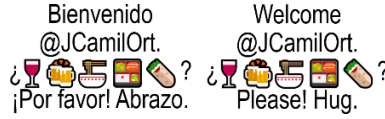

Fig. 1. Emoji use referring food consumption.

Taking into account these insights, we had two main challenges: create a knowledge base that can be used to analyze food mention behavior in narrative texts from social networks and propose a **new food mention identification algorithm** that recognizes the context of food-related words using different aspects of the publication to disambiguate it, like hashtags, user mentions, emojis and food n-grams with n >= 1 as well as non-food n-grams.

3.1 Knowledge base generation

In this section we present the result of the knowledge base generation, which is composed of 11 lists, namely:

— *Emoji list:* this list was constructed using as primary source the 11th version of the Unicode emoji characters and sequences from Unicode standard. This list has a total of 2620 emojis.

— *Food emoji list:* Felbo et al., [10] used the emoji prediction to find topics related to feeling in different domains. In our proposal we use a subset of the 95 emojis in the *emoji list*, named as the *food emoji list*. An extract of this list is presented in figure 2, where both the emoji and its meaning in English and Spanish, can be seen.

Emoji	Food	Spanish word
	Avocado	Aguacate
	Bread	Pan
	Chicken	Pollo
	Burger	Hamburguesa
	Ice cream	Helado

Fig. 2. Food emoji list sample.

— *What list:* to construct this list, the initial auto generated food list (see section 3) was manually reviewed, generating a new list which considers only unigrams, used on the food consumption context. It contains 776 words including their stem.
— *Where list:* this list allows identifying places, locations or spatial situations associated with food consumption. It contains 128 words.
— *Who list:* this list enables to identify people with whom food is shared using relationships and professions. It contains 112 words.
— *When list:* this list has moments, occasions and temporary situations in which people consume food. It contains 27 words.
— *Food stop word list:* it contains 178 popular sayings or expressions (non-food n-grams) frequently used on Twitter with food words, which do not correspond to the food consumption context. This list aims to be a filter to discard tweets.
— *Food list:* this list is composed of 441 n-grams with n > 1.
— *Food user mention list:* this list details 95 Twitter usernames associated with products, brands and places.
— *Food entity list:* food brand list with 812 elements.
— *Food hashtag list:* includes 450 hashtags related to food, that are typically used to refer to specific products or places.

According to the beforehand described lists, in the following section we present the proposed algorithm to identify food mentions in Twitter text.

3.2 Modelling

The proposed algorithm focuses on determining whether the tweet can be related to food context by text or by entity. In order to accomplish this, the main input of the algorithm is the tweet preprocessed text, which contains tokens, their stem and recognized entities. The algorithm is shown on figure 3 and explained next:

First, if the tweet only contains non-food n-gram, it is discarded and the algorithm finishes. Otherwise, the algorithm tries to determine a food context relationship by text or by entity:

- **By text:** for each token in a tweet, the algorithm checks if its stem belongs to the *what list*, if so, it validates the token only if it is a noun. If there are no more tokens to check, the algorithm determines a food context relationship by text.
- **By entity:** the algorithm determines a food context relationship by entity, if the tweet contains food context n-grams, brands, hashtags, mentions or emojis using the recognized entities from the preprocessed text.

Consequently, the algorithm assures a food context relationship only if, on the previous steps, at least one relationship or food mention was determined. In that case, the algorithm tries to identify context characteristics like places, people, or moments, using the *where, who* and *when lists*. Otherwise, the tweet is discarded. As a result, the algorithm stores five types of elements: (i) food n-grams, product or brand; (ii) place, whose identification is made through words, hashtags or mentions; (iii) people, with whom the food is consumed; (iv) moment of consumption (time of day, consumption time, day of the week, among others) and finally, (v) tweet publication time.
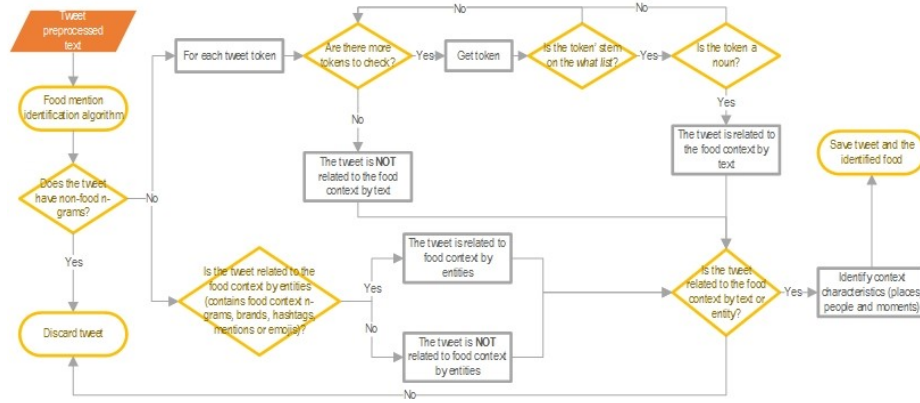

Fig. 3. Proposed food mention identification algorithm

3.3 Evaluation

To evaluate the proposed algorithm, an ETL (Extract, Transform, Load) system was designed and implemented using Big Data technologies. As shown in figure 4, Twitter

is used as data source, which is extracted using its public API[1] implementation, in Python[2].
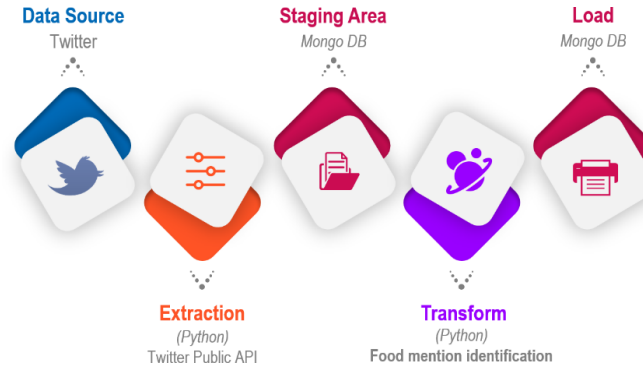

Fig. 4. ETL implementation design.

The extracted data are stored in a MongoDB[3] database, in the *Staging area*. Then, in the *transformation stage*, four steps take place:

1. *Data cleaning*: in this step, data is selected from the *Staging area* to be prepared, only if their geographic location corresponds to Colombia and their language to Spanish.
2. *Text preprocessing:* in this step, Twitter text is tokenized, tagged, parsed, stemmed and lemmatized using the spaCy's [4] Spanish processing pipeline. Additionally, the structured text (user mentions, hashtags and emojis - using the *emoji list* described on section 3.1.) is extracted and labeled accordingly.
3. *Named Entity Recognition (NER)*: here, the n-grams from the following lists are recognized and labeled as entities within the text: *food list, food stop word list, food user mention list, food hashtag list* and *food entity list*.
4. *Food mention identification*: in this last step, our algorithm (see section 3.2) is used to identify whether or not a tweet contains a food mention.

Finally, if the tweet contains a food mention, it is uploaded to Mongo DB in the *load stage*. It is worth mentioning that this exercise is part of a project named Digital Segmentation System from CAOBA [13], which, based on the information from Twitter, generates an approximation to the characteristics of the users who create the

---

1. About Twitter's APIs, https://help.twitter.com/en/rules-and-policies/twitter-api, last access date: 3rd May 2018, *we also provide companies, developers, and users with programmatic access to Twitter data through our APIs.*
2. Python, https://www.python.org/, last access date: 3rd May 2018, *Python is a programming language that lets you work quickly and integrate systems more effectively.*
3. MongoDB, https://www.mongodb.com/, last access date: 3rd May 2018, *building on the Best of Relational with the Innovations of NoSQL.*
4. spaCy, https://spacy.io/, last access date: 18th April 2018, *spaCy v2.0 Features new neural models for tagging, parsing and entity recognition.*

publications. These characteristics range from sociodemographic aspects, to socio-graphic attributes related to emotions, interests, and polarity. These variables will be considered in the next section.

Taking into account the ETL system, a case study was constructed with 1.3 million tweets extracted during fifteen days within the same month. In this period, our proposed algorithm identified 11,691 tweets that mentioned food, corresponding to 2 percent of the extracted tweets. A sample of the results obtained from the algorithm were manually evaluated to identify if the original tweet is actually related to the food context; as a result, a **precision** of 70 percent was obtained.

4   Results

The loaded tweets were classified depending on the type of the mentioned words. Our method manages to identify 1,310 different words, where 59 are mentioned 100 or more times. Table 1 shows the 20 most frequent words according to the time of day (breakfast: 5am-9am, lunch: 11am-2pm, snack: 10am/ 3pm-5pm and dinner: 6pm-9pm).

Table 1. Number of tweets according to the type of content entity. All words mark with * cannot be translated to English.

Breakfast		Lunch		Snack		Dinner	
Word	**%**	**Word**	**%**	**Word**	**%**	**Words**	**%**
cafe (coffee)	15,6%	cerveza (beer)	19,6%	cerveza (beer)	19,7%	cerveza (beer)	21,5%
cerveza (beer)	14,7%	torta (cake)	13,0%	torta (cake)	10,6%	comer (eat)	12,6%
torta (cake)	13,6%	almorzar (have lunch)	11,9%	comer (eat)	9,5%	cafe (coffee)	7,1%
comer (eat)	9,3%	comer (eat)	9,8%	cafe (coffee)	8,6%	pizza (pizza)	6,5%
bebida caliente (hot drink)	8,8%	cafe (coffee)	8,4%	aguacate (avocado)	6,6%	aguacate (avocado)	6,5%
pan (bread)	4,6%	pizza (pizza)	3,1%	vino (wine)	4,5%	torta (cake)	6,2%
tinto (black coffee)	4,5%	pollo (chicken)	3,1%	almorzar (have lunch)	3,7%	hamburguesa (burguer)	4,2%
arepa (*)	2,8%	pan (bread)	3,0%	pan (bread)	3,7%	pan (bread)	3,8%
carne (meat)	2,7%	tragar (swallow)	2,9%	pizza (pizza)	3,7%	tragar (swallow)	3,6%
tragar (swallow)	2,5%	carne (meat)	2,8%	coctel (coctel)	3,6%	jugar (play)	3,1%
chocolate (chocolate)	2,4%	coctel (coctel)	2,5%	tragar (swallow)	3,3%	coctel (coctel)	3,1%

almorzar (have lunch)	2,3%	arroz (rice)	2,5%	pinchar (*)	3,2%	chocolate (chocolate)	3,0%
pizza (pizza)	2,1%	hamburguesa (burger)	2,4%	chocolate (chocolate)	2,9%	pinchar (*)	2,8%
queso (cheese)	2,1%	chocolate (chocolate)	2,4%	bebida caliente (hot drink)	2,8%	queso (chesse)	2,5%
coctel (coctel)	2,1%	vino (wine)	2,4%	hamburguesa (burger)	2,6%	vino (wine)	2,5%
empanada (*)	2,1%	mango (mango)	2,3%	pollo (chicken)	2,5%	pana (friend)	2,5%
caballo (horse)	2,1%	pana (friend)	2,3%	carne (meat)	2,3%	empanada (*)	2,5%
aguacate (avocado)	1,9%	bebida caliente (hot drink)	2,0%	empanada (*)	2,2%	pollo (chicken)	2,4%
papaya (papaya)	1,9%	queso (cheese)	1,9%	queso (cheese)	2,1%	arepa (*)	1,9%
hamburguesa (burger)	1,8%	arepa (*)	1,8%	tinto (black coffee)	1,9%	carne (meat)	1,9%

In general, the word "cerveza" (beer) is the most frequent, almost at any time of the day; however, there is a group of words showing the consistency with a Colombian dietary routine to be mentioned at a specific time of day, such is the case of *bebida caliente (hot drink), pan (bread), queso (cheese), arepa (white corn cake) and chocolate(chocolate)* at breakfast; or *arroz (rice), pollo (chicken), pizza y carne (pizza and meat)* at lunchtime.

According to the previously established classification, it was found that, 33.108 times, a word was identified as food, product or brand; 1.324 times as a place, 1.426 times as a companion and 2.726 times as a consumption occasion. For the three last classifications, the most frequently mentioned words are presented in figure 5.

Fig. 5. Words cloud for where, who and when

Additionally, it is possible to know the behavior of users according to the day time in which they publish. Figure 6 shows a tendency to publish more tweets around noon and between 18 – 21 hours.
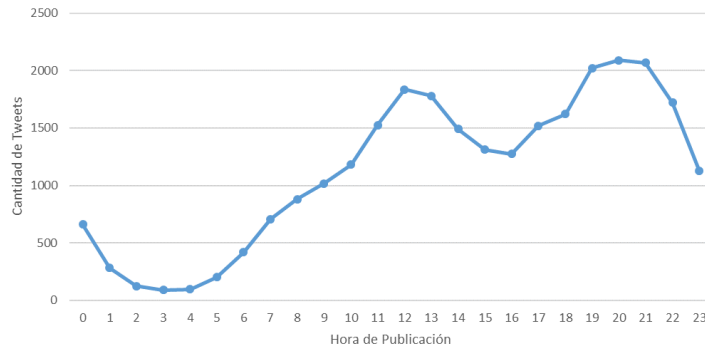
Fig. 6. Publication Frequency Distribution of tweets according to publication time

In relation to the sociographic variables, emotion is detected for 40 percent of the tweets and polarity for 86.1 percent. When performing the individual analysis of the most frequent words and their relationship with the emotion of the tweet, it is observed (see table 2) that words like cake, cocktail, hot drink, wine and avocado, have a high participation with tweets related to joy. On the other hand, words like dinner, pizza and chicken, have shares in sadness exceeding 20 percent of the tweets.

Table 2. Most frequent words according to emotions. All words mark with * cannot be translated to English.

Words / Emotion	Happiness 😄	Anger 😡	Fear 😰	Repulsion 🤮	Surprise 😮	Sadness 🙁	Total
cerveza (beer)	87,2%	2,5%	1,3%	1,1%	1,4%	6,5%	100,0%
torta (cake)	97,9%	0,2%	0,2%	0,0%	0,5%	1,2%	100,0%
comida (dinner)	57,7%	6,3%	2,8%	4,8%	4,4%	23,9%	100,0%
café (coffee)	72,6%	4,5%	2,9%	2,1%	3,2%	14,7%	100,0%
pizza (pizza)	65,8%	5,2%	3,6%	2,4%	2,7%	20,3%	100,0%
almorzar (lunch)	60,9%	7,6%	4,5%	1,4%	1,7%	23,9%	100,0%
coctel (cocktail)	93,2%	1,1%	2,2%	1,4%	0,4%	1,8%	100,0%
bebida caliente (hot drink)	91,1%	0,4%	1,9%	1,9%	0,7%	4,1%	100,0%
vino (wine)	93,6%	0,8%	1,1%	0,8%	0,4%	3,4%	100,0%
trago (*)	63,6%	5,7%	2,7%	2,7%	2,3%	23,0%	100,0%
chocolate (chocolate)	68,6%	5,7%	2,9%	1,4%	2,9%	18,6%	100,0%
aguacate (avocado)	92,9%	1,9%	1,0%	1,0%	0,0%	3,3%	100,0%
pollo (chicken)	65,3%	6,0%	3,3%	2,7%	1,3%	21,3%	100,0%
queso (cheese)	72,2%	3,5%	5,6%	2,1%	2,1%	14,6%	100,0%

hamburguesa (burger)	75,9%	3,0%	1,5%	3,0%	0,0%	16,5%	100,0%

4.1   Characterizing users by age and gender.

The following table (see table 3) shows the grouping by type of food and age groups of users who mention them, considering about 50 percent of the most mentioned words. It is observed that there is a trend in the 35-year old population, and more towards the mention of healthier foods, such as meat, cheese, chicken or the so-called "Natural Food" which in this text refers to as the usual or homemade food: rice, pasta, potato. The tendency to mention alcoholic beverages is strong in the Colombian tweets; however, it is much more pronounced in the population under 35 years.

Table 3. Food group versus age range.

Food group / age range	13-24	25-34	35 and more
Bebida alcohólica (alcoholic beverage)	36,6%	35,0%	26,4%
Bebidas (drinks)	13,0%	12,6%	16,5%
Carne, queso, pollo, queso (Meat, cheese, chicken, cheese)	9,5%	8,1%	10,0%
Comida natural (natural food)	15,9%	17,3%	18,5%
Comida rápida (fast food)	7,2%	5,1%	6,2%
Frutas, verduras (fruits, vegetables)	1,8%	1,1%	4,8%
Helados, postres (ice cream, desserts)	1,1%	0,9%	1,1%
Panes, tortas, arepas (breads, cakes, arepas)	14,9%	20,0%	16,5%
Total General	**100,0%**	**100,0%**	**100,0%**

When the differences at the gender level are observed, there is a tendency of men towards the mention of alcoholic beverages, such as wine, drink and beer; whereas in women, terms such as desserts, sweets, milkshakes and chocolates, are more frequent. That is, a pronounced tendency was found in women towards mentioning sweets; nevertheless, the mention of fruits by them, is also evident. These results are presented in the word clouds of figure 7.

Fig. 7. Food related words by gender (men and women)

The differences in the mentions of alcoholic beverages are also perceived at a so-cio-economic level, representing the greatest differences; such as fondness for beer into lower socio-economical levels, and the opposite for cocktail, a more expensive drink. Coffee and hot drinks also predominant in high strata. The following table (see table 4) presents the words showing the greatest differences between strata.

Table 4. Word distribution by socioeconomic level. All words mark with * cannot be translated to English.

Word / Socioeconomic level	High	Medium	Low
Cerveza (beer)	25,21%	32,09%	34,75%
Trago (shot)	3,55%	5,81%	9,62%
Pan (bread)	7,11%	4,89%	6,53%
Pizza (pizza)	5,72%	6,62%	8,32%
Empanada (*)	0,23%	0,11%	0,16%
Aguacate (avocado)	0,78%	8,05%	2,77%
Pollo (chicken)	5,03%	3,32%	6,20%
Pasta (pasta)	2,72%	1,94%	3,92%
Arroz (rice)	3,19%	2,60%	4,08%
Tinto (black coffee)	0,88%	0,62%	0,98%
Arepa (*)	3,83%	2,55%	1,79%
Coctel (cocktail)	5,12%	4,59%	2,45%
Carne (meat)	5,54%	3,09%	2,61%
Jugo (juice)	3,60%	3,43%	6,04%
Café (coffee)	3,37%	1,81%	1,96%
Bebida caliente (hot drink)	8,59%	3,39%	1,14%
Torta (cake)	15,51%	15,10%	6,69%
Total	**100,0%**	**100,0%**	**100,0%**

5   Discussion and future work

Food preferences expressed in social networks as Twitter become a valuable source of information for making decisions about consumer centralization strategies. There-fore, knowing tendencies of publishing, interests and behaviors based on comments published by users allows identifying pertinence and strength of marketing strategies.

Through a case study, it is shown that Twitter information provides some elements that allow a global analysis of preferences in foods, products or brands, based on the mentions made by users in the network. Despite founding just a 2% of messages related to food, there is a significant number of users, which would significantly exceed ap-proximations made by other methodologies, such as specialized surveys. The advantage

of having continuous information collection also enables a significant increase in the volume of users that can be identified over time, as well as the identification of patterns or changes in behavior, constituting a very relevant aspect.

One of the most valuable elements of the exercise is the generation of the knowledge base, which must be adjusted to ensure that the products of interest (including those of competitors) are at the base; in turn, more specific relationships can be established about users' opinions or emotions about them. An advantage of the way in which the algorithm was implemented is the possibility of making these adjustments without major difficulties. This would create new sources of unstructured open data, allowing other systems to feed from their knowledge bases, such as systems of health, marketing or others [15].

Despite the remarkable advantages, it is important to note that the algorithm's accuracy is 70 %, a value associated mainly with trying to build a knowledge base for such a broad domain. This behavior affects the results, generating erroneous interpretations; however, as mentioned before, if this algorithm was applied to a more specific domain, its performance would increase.

To estimate the magnitude of interpretation errors, it will be necessary to deepen in a content analysis where the intentionality is verified directly in the tweet texts. This kind of analysis requires a huge amount of time for its completion, which exceeds the initial objectives and scope of this research. However, it is proposed as a future work.

6   Acknowledgements

References

1. Vidal, L., Ares, G., Machín, L., & Jaeger, S. R. (2015). Using Twitter data for food-related consumer research: A case study on "what people say when tweeting about different eating situations". Food Quality and Preference, 45, 58-69.
2. Nielsen Company. (2014). Advertising and audiences: State of the media May 2014. Retrieved from http://www.nielsen.com/content/dam/nielsenglobal/jp/docs/report/2014/Nielsen_Advertising_and_ Audiences Report-FINAL.pdf
3. Janasz, T., Koschmider, A., Born, M., & Uhl, A. (2016). Digital Capability Framework: A Toolset to Become a Digital Enterprise. In Digital Enterprise Transformation (pp. 51-84). Routledge.

4. Uhl, A., MacGillavry, K., & Diallo, A. (2016). Digital Transformation at DHL Freight: The Case of a Global Logistics Provider. In Digital Enterprise Transformation (pp. 287-302). Routledge.

5. Abbar, S., Mejova, Y., & Weber, I. (2015, April). You tweet what you eat: Studying food consumption through twitter. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 3197-3206). ACM.

6. Prier, K. W., Smith, M. S., Giraud-Carrier, C., & Hanson, C. L. (2011, March). Identifying health-related topics on twitter. In International conference on social computing, behavioral-cultural modeling, and prediction (pp. 18-25). Springer, Berlin, Heidelberg.

7. Dredze, M., Paul, M. J., Bergsma, S., & Tran, H. (2013, June). Carmen: A twitter geolocation system with applications to public health. In AAAI workshop on expanding the boundaries of health informatics using AI (HIAI) (Vol. 23, p. 45).

8. Moreno-Sandoval L., Beltrán-Herrera P., Vargas-Cruz J., Sánchez-Barriga C., Pomares-Quimbaya A., Alvarado-Valencia J. and García-Díaz J. (2017). CSL: A Combined Spanish Lexicon - Resource for Polarity Classification and Sentiment Analysis.In Proceedings of the 19th International Conference on Enterprise Information Systems - Volume 1: ICEIS, ISBN 978-989-758-247-9, pages 288-295. DOI: 10.5220/0006336402880295

9. Moreno-Sandoval L., Mendoza-Molina J., Puertas E., Duque-Marín A., Pomares-Quimbaya A. and Alvarado-Valencia J. (2018). Age Classification from Spanish Tweets - The Variable Age Analyzed by using Linear Classifiers.In Proceedings of the 20th International Conference on Enterprise Information Systems - Volume 1: ICEIS, ISBN 978-989-758-298-1, pages 275-281. DOI: 10.5220/0006811102750281

10. Felbo, B., Mislove, A., Sogaard, A., Rahwan, I., & Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. arXiv preprint arXiv:1708.00524.

11. Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011, October). Our twitter profiles, our selves: Predicting personality with twitter. In Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (Social-Com), 2011 IEEE Third International Conference on (pp. 180-185). IEEE.

12. Krishnamurthy, B., & Wills, C. E. (2008, August). Characterizing privacy in online social networks. In Proceedings of the first workshop on Online social networks (pp. 37-42). ACM.

13. Vargas-Cruz, J., Pomares-Quimbaya, A., Alvarado-Valencia, J., Quintero-Cadavid, J., & Palacio-Correa, J. (2017). Desarrollo de un Sistema de Segmentación y Perfilamiento Digital. Procesamiento Del Lenguaje Natural, 59, 163-166. Recuperado de http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5511

14. Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Stříteský, V., & Holzinger, A. (2015). Reprint of: Computational approaches for mining user's opinions on the Web 2.0. Information Processing & Management, 51(4), 510-519.

15. Valdez, A. C., Ziefle, M., Verbert, K., Felfernig, A., & Holzinger, A. (2016). Recommender systems for health informatics: State-of-the-art and future perspectives. In Machine Learning for Health Informatics (pp. 391-414). Springer, Cham.

16. Puertas Del Castillo, E., Alvarado Valencia, J., & Pomares Quimbaya, A. (2017). Constructor automático de modelos de dominios sin corpus preexistente. Procesamiento Del Lenguaje Natural, 59, 129-132. Recuperado de http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5503