



ACTIVITY OF BRAZILIAN TOURISM AGENCIES IN SOCIAL MEDIA: AN ANALYSIS USING NATURAL LANGUAGE PROCESSING

Danillo Magno Duarte Guedes
 <https://orcid.org/0009-0006-9091-3824>
danillomda@hotmail.com

Universidade federal de Minas Gerais (UFMG)
Belo Horizonte, MG, Brasil

Marlusa de Sevilha Gosling
 <http://lattes.cnpq.br/9503365193492380> –  <https://orcid.org/0000-0002-7674-2866>
mg.ufmg@gmail.com

Universidade federal de Minas Gerais (UFMG)
Belo Horizonte, MG, Brasil

ABSTRACT

This paper aims to analyze the activity of two Brazilian tourism agencies in social media and the online behavior of their consumers. The research used Natural language processing resources supported by sentiment and content analysis techniques. The main results show a prevalence of positive comments on the companies' pages, and companies are more responsive to users on pages with a higher number of positive comments. There is also a tendency towards more significant company interaction with user comments that express positive emotions.

Keywords: Content marketing. Social media. User-generated content. Natural language processing. Sentiment analysis.

ATIVIDADE DE AGÊNCIAS BRASILEIRAS DE TURISMO NAS REDES SOCIAIS: UMA ANÁLISE UTILIZANDO PROCESSAMENTO DE LINGUAGEM NATURAL

RESUMO

Este trabalho buscou analisar a atividade de duas agências brasileiras de turismo nas mídias sociais e o comportamento *online* de seus consumidores. A pesquisa utilizou recursos de processamento de linguagem natural, apoiados pelas técnicas de análise de sentimento e análise de conteúdo. Os principais resultados indicaram que há prevalência de comentários positivos nas páginas das empresas, sendo que as empresas são mais responsivas aos usuários em páginas que possuem um maior número de comentários positivos de seus clientes. Apontou-se, ainda, uma tendência de maior interação por parte das empresas com comentários de usuários que expressam sentimentos positivos.

Palavras-chave: Marketing de conteúdo. Mídias sociais. Conteúdo gerado pelo usuário. Processamento de linguagem natural. Análise de sentimento.

DOI <http://dx.doi.org/10.1590/1981-5344/25280>

Recebido em: 03/04/2023.

Aceito em: 11/05/2023.

1. Introduction

When discussing marketing in the current times, it is indispensable to mention the influence and impacts of social media on many market segmentations, for its rise and mass adoption have been providing excellent business opportunities for companies in the marketing field. Evidence of this tendency is the increased relevance of digital marketing in that area of research and testing since, in addition to using extraordinarily efficient and low-cost tools for advertising campaigns, social media has become an important communication channel between consumers and companies in recent years (EVANGELISTA; PADILHA, 2014).

In this context, the feedback generated by the comments can represent an influencing factor for consumers' behavior and their purchasing decisions. The study "The 2012 Traveler" (THINK..., 2012), carried out by Google and the Ipsos MediaCT institute, reveals that when people are willing to plan their trip, approximately 37% of leisure travelers use the comment sections as sources, looking for reviews and opinions. According to Ye, Law, and Gu (2009), online reviews written by hotel consumers have an important impact on hotel room sales. Therefore managers should seriously consider those reviews, especially the ones posted on third-party websites.

On the other hand, the way companies react to comments can also be an essential factor. As Pantelidis (2010) describes, restaurant managers who respond appropriately to comments on electronic forums can turn dissatisfied consumers into loyal customers.

In general, the content of comments in company posts on social media presents unstructured information that, when processed through ratings, can be used as a support tool for performance indicators. According to Miranda and Sassi (2014), sentiment analysis can be used as a support tool to enrich the assessment of consumer satisfaction. According to Stich, Emonts-Holley, and Senderek (2015), researchers can use the technique to assess the "consumer experience."

In addition to sentiment analysis, natural language processing techniques can provide information from the content of consumer comments that leads to the detection of defects and opportunities for improvements in the product or service in question (MOGHADDAM; ESTHER, 2010).

The growth of social media like Instagram and Facebook, which encourage the sharing of photos by their users, can represent an opportunity for sectors such as tourism, which can represent one of the activities that most use images to promote themselves and to attract consumers, primarily through sharing photos and experiences on social media (SANTOS *et al.*, 2017).

Given the arguments and studies presented, as well as the growth of techniques that enable the analysis of a large amount of data, the volume of academic studies on the content of comments in posts by Brazilian companies on social media is relatively low (e. g. AMARAL *et al.*, 2017; COELHO; GOSLING, 2015; EVANGELISTA; PADILHA, 2014). Also, most studies observe the need to use the available tools for sentiment analysis or other natural language processing techniques. Therefore, this article aims to analyze the social media activity of Brazilian consumers and companies in the tourism sector.

The information resulting from this type of research can contribute not only to more in-depth studies on the behavior of consumers and companies on social media but also to companies that want to increase their presence on social networks, seeking to leverage the visibility of their brand.

2 Theoretical reference

2.1 Content marketing

According to Baltes (2015), content marketing has become the key to marketing campaigns' success and the essential tool in digital marketing. Visual resources, such as images, videos, and others, are often used to create content for social media. Vicente (2016) points out that posts with images have advantages in promoting engagement over other types of posts. In addition, the author highlights the power of storytelling and narratives in the photo captions to significantly impact the consumer.

According to Forouzandeh, Soltanpanah and Sheikahmadi (2014), content marketing has a significant advantage compared to other types of marketing on social media. Users are not directly and suggestively introduced to this type of marketing. Instead, they are involved in another matter and are provided with helpful information they do not consider commercial, making them trust the content provider. Establishing a feeling of trust promotes users' loyalty. Lima (2014) considers this loyalty a consequence of an engagement with

the brand. Therefore, content marketing and consumer engagement efforts are directly related.

The social media platform Facebook is an adequate basis for content marketing, through which different types of content can be easily presented and quickly distributed among users. Also, marketing professionals can combine Content marketing with other forms of marketing. By developing and promoting knowledge and information among users, content marketing creates a feeling of trust toward purchasing goods due to the content presented. In addition, this increase in trust among users expands the sale of goods. Content marketing is the knowledge to absorb customers indirectly. The results of the study by Forouzandeh, Soltanpanah and Sheikahmadi (2014) confirm such benefits of content marketing, as it can present various goods to users.

2.2 Social media in the tourism sector

For Zeng and Gerritsen (2014), social media plays an increasingly important role in several aspects of tourism, especially regarding the search for information and decision-making behaviors (FOTIS; BUHALIS; ROSSIDES, 2012), tourism promotion (BRADBURY, 2011) and focusing on best practices for interacting with consumers through social media channels (e. g. social sharing of vacation experiences). Many countries regard social media as essential to promote their tourism industries.

Leung *et al.* (2013) point out that several academics noted the ability of social media to help tourism and hospitality companies to engage potential customers, increase their online presence and thus lead to higher online revenues. For Agnihotri *et al.* (2016), the responsiveness acquired from informative communication with customers on social media positively correlates with consumer satisfaction. Gallagher and Ransbotham (2010) highlight that social media makes it easier for companies to have a firm-to-consumer dialogue, strengthening consumer-to-firm and firm-to-consumer communication.

The authors also concluded that consumers generally used social media during the research phase of their travel planning process. In addition, reliability is a critical antecedent in determining their decision about using the information on social media. Finally, the article discusses the applications of social media in five main functions (promotion, product distribution, communication,

management, and research). Based on the research results, social media is an important strategic tool in managing tourism and hospitality – particularly in promotion, business management, and research functions.

2.3 Sentiment analysis

The sentiment analysis technique seeks to create structured knowledge that a support system or decision-maker can use (ARAÚJO; BENEVENUTO; RIBEIRO, 2013). Specifically, in marketing and Customer Relationship Management (CRM), sentiment analysis aims to detect favorable or unfavorable opinions about products and services by using a large amount of virtual data in text format collected from social media, such as social networks, website recommendations, forums, blogs, and other sources (MORENO, 2015).

Sentiment Analysis usually classifies emotions into two categories: positive versus negative, or into three categories, considering neutral comments and a polarity score (SHARDA *et al.*, 2014) or even an opinion score (PANG; LEE, 2008).

According to Araújo, Benevenuto, and Ribeiro (2013), there are currently two main approaches to sentiment analysis of textual productions. The first one, the supervised approach, is based on machine learning concepts, which start from defining characteristics that allow one to distinguish between sentences with different emotions by training a model with previously labeled sentences. The model then can identify the emotion in previously unknown sentences. The second approach, the unsupervised one, does not rely on machine learning model training and, in general, is based on lexical methods for treating emotions that involve calculating the polarity of a text based on the semantic orientation of the words it contains. Although previously labeled data is not necessary to carry out the training, its efficiency is directly related to the generalization of the vocabulary used, considering the various existing contexts.

Several sentiment analysis methods, supervised and unsupervised, are available in the literature. Ribeiro *et al.* (2016) compare the predictive capacity of 18 different methods. To perform the comparison, The use of datasets containing texts that were earlier manually labeled with emotions and separated into two or three categories was necessary. The metrics used are accuracy, precision, recall, and the F1 score (GONÇALVES *et al.*, 2013). The authors point out that, although the results have identified some methods considered among

the best for different datasets, the overall prediction performance still left much room for improvement. More importantly, the predictive performance of the methods varies widely across data sets. Consequently, the level of agreement between the methods greatly varies when analyzing the same body of text. As expected, the methods' two-class prediction capacity is considerably greater than their three-class prediction capacity.

In addition to the standard methods used and documented in the literature, specific methods are available, such as IBM Watson and Microsoft Text Analytics, which are implementations mainly focused on commercial applications. However, they have resources available for testing, through which their users are free to use tools with certain limitations (AGUIAR *et al.*, 2018).

As Araújo *et al.* (2016) explain, most of those resources are available only in English, considering that this language dominates the content provided. However, some efforts have been present to develop emotion techniques in other languages, although there is little knowledge about the performance and the absolute need or feasibility of developing these solutions. Therefore, the authors propose using specific state-of-the-art methods for analyzing emotions in nine languages. To that end, they use data previously labeled in each language and a simple automatic translation into English and develop a methodology to compare and validate the results. However, there is still little research in the literature regarding the metrics of accuracy and precision of other existing methods, such as Watson Natural Language Understanding (NLU), in detecting the emotions in texts written in the Portuguese language (*e. g.*, AGUIAR *et al.*, 2018).

According to Gonzales and Lima (2003, p. 3)

Natural language processing (NLP) computationally handles the different aspects of human communication, such as sounds, words, sentences, and speeches, considering formats and references, structures and meanings, contexts and uses. In a vast sense, NLP aims to make the computer communicate in human language, not always necessarily at all levels of understanding and generation of sounds, words, sentences, and speeches.

According to Pang and Lee (2008), companies have paid more attention to collecting all information published about their products, services, reputation, and customers. In addition, consumers are also interested in knowing relevant information about products they may consume or what is being said about them

(THINK..., 2012; YE; LAW; GU, 2009). These facts point to a constant implementation by companies seeking information that provides competitive advantages of techniques brought by advances in information technology, especially in the digital environment.

2.4 Sentiment analysis in tourism

Table 1 - Studies related to sentiment analysis in the context of tourism

Authors	Results
Dwivedi <i>et al.</i> (2007)	The Authors indicate an imminent loss of control over the network's information regarding companies and industries.
Thevenot (2007)	The study concluded that if companies would not properly monitor online comments, the industry would bear the consequences, as blogs generate positive and negative impacts.
Blair-Goldensohn <i>et al.</i> (2008)	The authors utilized Google Maps data to construct the sentiment analysis tool to analyze consumers' emotions regarding hotels, stores, and restaurants, determining the value of the polarities (positive/negative). In addition, the system summarized emotions concerning different aspects of the services provided, such as price and environment.
Pekar and Ou (2008)	The authors use Sentiment analysis to evaluate 268 customer reviews of large hotels published on the 22epinions.com website. The authors use resources such as food, room service, facilities, and price to analyze automatically the emotions expressed regarding each hotel.
Barcelos <i>et al.</i> (2015)	The results demonstrate how the categorization by themes and data visualization in graphs can condense the information available on a social media platform specialized in tourism (TripAdvisor).
Bakhshi <i>et al.</i> (2015)	The study identifies that the most active and regular members of the community are the ones that most contribute to the good quality of the comments and that the most extended comments are more likely to be popular, receiving stars and votes. Also, consumers tend to see useful comments first.
Geetha and Sinha (2017)	The authors identify consistency between the customer ratings and their real emotions regarding hotels of both categories of, premium and budget. The polarity of customer emotions explains the significant variation in customer ratings in both hotel categories.

Source: prepared by the authors (2019) based on Moreno (2015).

According to Moreno (2015), there needs to be more support in the literature concerning applying sentiment analysis techniques in tourism. Table 1

presents the literary studies carried out over the years related to applying sentiment analysis techniques in the context of the tourism sector.

The first publications made in the area appeared in 2007. They address the unexpected influence of social media on companies and the tourism industry (ZENG; GERRITSEN, 2014) and the imminent loss of control over the information available on the network regarding both companies and the industry (DWIVEDI *et al.*, 2007). With emphasis, Thevenot (2007) already indicated that if companies would not properly monitor online comments, the industry would have to deal with the consequences since blogs also generate negative impacts (THEVENOT, 2007).

3. Methodological procedures

Table 2 - Research Strategy

Specific objective	Analyzed Units	Data collection	Data Analysis
Analyze the emotions expressed in the comments on the posts	Company A (Facebook: 7536 comments on 221 posts from 04/2018 to 04/2019.	Manual Collection. Automated data structuring using the Python language.	Tools: Sentiment Analysis - iFeel and IBM Watson NLU Metrics: P/NG, P/Total, NG/Total, P/(P + N), NG/(P+N)
Investigate whether there is a difference in the emotion expressed in posts from different social media.	Instagram: 7907 comments on 98 posts from 07/2018 to 04/2019).		Tools: T-Student Test Metrics: average of sentiment polarity
Investigate the degree of companies' responsiveness regarding complaints, suggestions, and other manifestations of their users.	Company B (Facebook: 8344 comments on 46 posts from 11/2018 to 04/2019.		Tools: Response Index Metrics: Responses to negative comments/total responses
Look for signs of service aspects with potential for improvement.	Instagram: 7917 comments on 76 posts from 01/2019 to 04/2019).		Content analysis, Natural Language Processing (spaCY Library and IBM Watson NLU)

Source: prepared by the authors (2019).

Table 2 presents the research strategy. We can classify this research as exploratory. Malhotra (2001, p. 106) defines exploratory research as "a type of research whose main objective is to provide criteria on the problem situation faced by the researcher and their understanding". In addition, the research used text mining techniques, classified by Hearst (1999) as exploratory data analysis. On the other hand, the data analysis process considered mainly quantitative factors, such as the rate of positive/negative comments. However, the analysis also benefited from the qualitative aspects of the data present in the content of the texts. Thus, the research followed a qualitative and quantitative approach.

3.1 Data collection and structuring

We carried out this research by collecting unstructured data. The sources for collection are the data contained in posts of two Brazilian travel agencies with great relevance in the market (LUCAS, c2013). The choice of companies was due to their online presence. Travel agencies with large numbers of followers on social media were selected to provide a greater volume of unstructured data for the survey. There is no delimitation on socio-demographic factors in the selection of post comments.

3.2 Data Analysis

3.2.1 Polarity of comments

To analyze the emotions expressed in the comments present in the posts, the iFeel tool, provided by the Department of Computer Science at the Federal University of Minas Gerais (ARAÚJO *et al.*, 2014), was primarily used. The tool, available in 60 languages, performs the sentiment analysis of texts and files using 18 methods available in the literature to achieve greater coverage and accuracy. First, the text is translated into English using the Yandex Application Programming Interface (API) to reach this mark and then analyzed. According to Reis *et al.* (2015), this approach proved feasible since the translation of databases into other languages does not significantly interfere with the accuracy or the range of methods for sentiment analysis.

According to Araújo *et al.* (2016), the most accurate method of detecting emotions using machine translation is Sentistrength. This method uses a lexicon

dictionary labeled by humans and enhanced by machine learning. However, methods like Umigon and Vader also performed well in different databases. In addition, according to Ribeiro *et al.* (2016), the Sentistrength method tends to classify many sentences as neutral, which can decrease its accuracy in analyses considering three classes (positive, negative, and neutral). Thus, this research considers the results of the Sentistrength and Umigon methods. The results for each comment were reintroduced in the columns next to it in the spreadsheets initially set up.

The comments were also analyzed to complement the research and make comparisons using a script programmed in Python language, which uses the sentiment analysis feature available in the IBM Watson API. (Watson Natural Language Understanding), due to its ability to perform analysis natively in Portuguese (SOUSA, 2017). In addition, there is a limitation in the iFeel tool that prevents the analysis of sentences with more than 300 characters. Therefore, comments that overcame this barrier were also analyzed separately through IBM Watson, employing a script that goes through the spreadsheets containing the iFeel results, identifies the comments with more than 300 characters, analyzes them, and saves the results. We carried out a test of the method's ability to identify the emotions present in the texts, according to the metrics used by Gonçalves *et al.* (2013), to check the feasibility of using IBM Watson in its current state of development. For this, a database of comments written in Portuguese on social media Twitter was used, manually labeled by the MiningBR group (AGUIAR *et al.*, 2018). The analysis results are at the end of the discussion and results section.

A script developed in Python that runs through each line of the spreadsheet counted and calculated the metrics. Due to its structure, we can associate comments with the respective post. We consider all comments except those posted on the official companies' websites. The analysis metrics chosen were: the ratio between positive and negative comments (P / NG), the ratio between positive and total ($P / Total$), and between negative and total ($NG / Total$). Since the measures of accuracy and precision of the methods present in the literature are higher in the analysis of two classes (positive and negative), we also present the metrics omitting the number of neutral comments, the ratio of each class by the sum of the two classes ($P / (P + NG)$ and $N / (P + NG)$).

By observing the metrics, we analyzed the differences between the results of the two social media of each company. A T-Student statistical test ensured the results' significance, comparing the recorded polarity averages with a 95% confidence interval and assuming different variances between the samples. For this test, we transform the results into values ranging from -1 (negative) to 1 (positive). We used the Levene test (SCHULTZ, 1985) to decide on the assumed equality or variance difference between samples. In the case of company A, since the observed Facebook comments are far older than the first Instagram posts, the T-Student test was carried out on a sample where all comments on both media are from between 09/09/2018 and 03/25/2019, under the premise that possible changes in the company's services or events over time can impact the average polarity of comments.

3.2.2 Companies responsiveness

An automated search in each spreadsheet analyzed the companies' responsiveness through the development of a script in the Python language (whose operation description is present in the Appendix) using the libraries: xlutils, xlrd, and xlwt and, to extract the entities present in the text, the SpaCy library for Natural Language Processing, which allows a process of separating the sentences into their respective grammatical and syntactic classes, in addition to extracting the entities present in the text. In 2015, a survey by Emory University and Yahoo! Labs showed that spaCy offered the fastest syntactic parser in the world and that its accuracy was among the best available (CHOI; TETREULT; STENT, 2015).

In this way, each comment answered by the company, even those present in the responses of another comment, is associated with a company response. Finally, this newly created spreadsheet was scrolled down to search, based on the total of answered comments, the same metrics presented for the polarity of the total comments.

3.2.3 Aspects of improvement

The search for improvement aspects in the companies' service started with extracting nouns and their related adjectives and keywords in the text

content, using the SpaCy library and the IBM Watson API NLU module available in Python.

The developed script ran through an aggregated spreadsheet containing comments on the pages of the two social media of each company, extracting a list of nouns from each comment and, if available, a list containing the associated adjectives. Those elements are inserted in a second spreadsheet so that each line has a noun and an adjective. After that process, the script goes through the second spreadsheet, counting the incidence of nouns and noun-adjective pairs. Thus, it was possible to create a table with the spreadsheet data containing the nouns of the highest incidence and their recurrently associated adjectives in order of highest incidence. Nouns whose meanings are too broad or do not contain the aspects of improvement have been disregarded (e. g., "time," "day," and the name of the companies). In order to search for non-identified relevant keywords with the search for nouns and adjectives, the keyword identification feature of Watson NLU scrolled through the spreadsheets containing the results of this tool's sentiment analysis.

To complement the insights generated from the most relevant nouns, an analysis of the individual content of a portion of the comments considered negative was made, according to the procedures proposed by Bardin (2011).

In the pre-analysis stage, we defined that the comments would be analyzed assuming users reported some aspects of the service negatively. The objective was to find them. First, a spreadsheet for each company was separated, containing user comments classified as negative by at least one of the methods. The choice of documents followed the rule of representativeness. Next, we chose a portion of the negative comments for analysis. The previously identified nouns served as a guide in choosing comments so that, from the 15 most recurring nouns, were separate those with more specific meanings and related to aspects of service in the universe of tourism. From these words, we separate the five with the highest incidence.

From that, we defined using Excel to search for comments containing these words. The following script carried out the number of searches: In words or groups of words that totaled more than 200 incidences, it performed 100 searches on the spreadsheet for comments containing the word; in words whose incidence was less than 100, it performed 50 searches; when the incidence was

less than 50, it searches all comments containing the word in question. Then, it classifies the comments into more minor reports such as "Ticket canceled, route or date changed," "Problem in the post-sale," or "undue charge." It then counts the incidence of these reports. The results of the inferences from these data are available in section 4.3.

4. Results and discussion

When observing the companies' posts, a pattern is identified, mainly in company A, where resources such as images, videos, and informative descriptions of places and experiences are used, as well as inspiring texts and narratives, in order to awaken the consumers' desire (VICENTE, 2016) and to get them to seek the company's services. Thus, companies use content marketing strategies since users are not introduced directly and suggestively to the company's services, as Forouzandeh, Soltanpanah and Sheikahmadi (2014) explain. However, on the other hand, company B also uses many posts that resemble traditional advertisements, containing promotions and exposing its services.

4.1 Sentiment polarity

Table 3 illustrates the size of the samples in which the polarity analysis was performed, in addition to data such as the total of posts and the percentage of company comments concerning the total. It is important to emphasize that, for this stage, the research does not consider the comments of the analyzed companies.

Table 3 - Size of samples for calculating the sentiment polarity

	Company A		Company B	
	Instagram	Facebook	Instagram	Facebook
Total posts	90	216	76	44
Total followers comments	6677	5631	6498	5365
The average number of comments per post	74,19	26,06944	85,5	121,19
Percentage of company comments	15,55%	25,28%	17,92%	35,7%

Source: prepared by the authors (2019).



Concerning company A, the discrepancy observed in the average amount of comments per post between the two media indicates that the posts on the Instagram page achieve greater success in promoting engagement with their customers. Thus, this platform may improve performance in promoting digital marketing (ARAÚJO, 2015). Furthermore, the fact may be more likely when considering the differences in followers of the two media, since the company's Facebook page, at the time of the survey, had more than ten times the number of followers on the Instagram page (12,550,820 and 954,000 followers, respectively). However, regarding company B, the opposite situation is observed. Therefore, one possible hypothesis is that the posting style of company A, which uses more resources such as images and narratives in the caption, is more successful in promoting engagement on the social media Instagram than on Facebook, even though it has the network used as a basis by Vicente (2016) to point out the advantages of this type of posting. However, it is important to note that such findings can also simply reflect that companies have different strategies and ways of investing on each social media.

Table 4 - Sentiment polarity of comments on company A's page (Instagram)

	Method	Positive	Negative	Neutral	P/NG	P/Total	NG/Total	P/(P+NG)	NG/(P+NG)
Total	SentiSTR	2142	613	3922	3,494	0,320	0,0918	0,7774	0,2225
	Watson	3779	1340	1741	2,820	0,5508	0,1953	0,73823	0,26177
	Umigon	1525	843	4309	1,809	0,2283	0,1262	0,644	0,3559
Average per post	SentiSTR	23,8	6,81111	43,57	5,430	0,299	0,1311	0,7222	0,2777
	Watson	41,988	14,8888	19,34	4,624	0,530	0,2435	0,70105	0,2989
	Umigon	16,944	9,36666	47,87	2,694	0,216	0,1763	0,5814	0,4074

Source: prepared by the authors (2019).

As shown in Tables 4 and 5, positive comments are prevalent over negative ones. As expected, there is a large discrepancy between the results of the methods used, as reported by Ribeiro *et al.* (2016). This fact, along with the high incidence of comments classified as negative, indicates plenty of room for improvement in the prediction of methods.

Table 5 - Sentiment polarity of comments on company A's page (Facebook)

	Method (+NLU)	Positive	Negative	Neutral	P/NG	P/Total	NG/Total	P/(P+NG)	NG/(P+N)
Total	SentiSTR	1384	418	3829	3,311	0,24578	0,074232	0,768	0,2319
	Watson	2133	1000	2379	2,133	0,386974	0,181422	0,6808	0,3191
	Umigon	884	528	4219	1,6742	0,15698	0,093767	0,626	0,3739
Average per post	SentiSTR	6,4074	1,9351	17,72	3,7385	0,25559	0,069246	0,7064	0,1916
	Watson	9,7397	4,5662	10,86	3,43348	0,373987	0,162645	0,6468	0,27552
	Umigon	4,0925	2,4444	19,53	2,0816	0,16741	0,085797	0,5678	0,2979

Source: prepared by the authors (2019).

Both on Instagram and Facebook, positive comments are prevalent over negative ones. These results align with the results of Coelho and Gosling's (2015) work in the context of hotel reviews, which shows a tendency for those who demonstrate engagement in social media regarding a product/service to have a positive perception of it. In addition, generally, there is no apparent difference in the proportions of negative and positive comments between the two media. However, there is a slightly higher incidence of negative comments on the Facebook page, which may indicate a greater tendency for users of this media to express negative feelings in their comments. To test the significance of this difference, the results of the T-Student test of the polarity mean are available (Table 6), considering the same period and assuming different variances, as well as the result of the Levene test that supports the choice of the line of analysis. In this case, we consider hypotheses H0 (there is no difference between the posts on the Facebook and Instagram pages of company A regarding sentiment) and H1 (there is a difference between the posts on the Facebook and Instagram pages of company A regarding sentiment).

Since the value of P ($T \leq t$) is less than 0.05, and t Stat exceeds the value t Critical, we reject H0 and can affirm that there are differences in the average sentiment polarity on the posts. This finding may reinforce the hypothesis that Instagram has a more significant potential for success in obtaining the benefits that Vicente (2016) pointed out of company A's posting style.



Table 6 - Difference between polarity averages on company A pages

	Sentistrength Method		Umigon Method	
	Facebook	Instagram	Facebook	Instagram
Mean	0,086333	0,210058	-0,01281	0,08747
Variance	0,340118	0,357453	0,292697	0,335977
Observations	3278	6184	3279	6185
P-value:	0,0000		0,0000	
Df	6823		7090	
† Stat	-9,73396		-8,36859	
P(T<=†) one-tail	1,51E-22		3,48E-17	
† Critical one-tail	1,645077		1,645069	
P(T<=†) two-tail	3,01E-22		6,96E-17	
† Critical two-tail	1,960312		1,960299	

Source: prepared by the authors (2019).

As seen in Tables 7 and 8, there are higher rates of positive comments with the comments on the company A page. In contrast to the previous case, on the Company B pages, there is a higher incidence of negative comments on the Instagram page over the Facebook page. This finding contributes in favor of formulating hypotheses contrary to the one in which users of Instagram media have a greater tendency to comment positively.

Table 7 - Sentiment Polarity of comments on company B's Instagram page

	Method (+NLU)	Positive	Negative	Neutra	P/NG	P/Tota	NG/Tota	P/(P+NG)	NG/(P+N)
Total	SentiSTR	2033	283	4182	7,183	0,312	0,0435	0,8778	0,122193
	Watson	1651	437	1563	0	0,452	0,1196	0,79070	0,209291
	Umigon	1168	397	4933	2,942	0,179	0,061	0,74632	0,253674
Average per post	SentiSTR	26,75	3,72368	55,02	10,47	0,314	0,0539	0,85873	0,128108
	Watson	9,7397	4,56621	10,86	3,433	0,373	0,16264	0,64684	0,275526
	Umigon	15,368	5,22368	64,90	3,718	0,196	0,0772	0,7153	0,271539

Source: prepared by the authors (2019).

Table 8 - Sentiment polarity of comments on company B's Facebook page

	Method (+NLU)	Positive	Negative	Neutral	P/NG	P/Total	NG/Total	P/(P+NG)	NG/(P+N)
Total	SentiSTR	2038	156	3171	13,06	0,379	0,029	0,9288	0,0711
	Watson	3178	362	1727	8,779	0,603	0,0687	0,89774	0,10226

Average per post	Umigon	1580	299	3486	5,284	0,294	0,0557	0,8408	0,15912
	SentiSTR	46,318	3,5454	72,06	12,16	0,362	0,0303	0,9199	0,08009
	Watson	72,227	8,22727	39,25	11,70076	0,580	0,0737	0,88588	0,114111
	Umigon	35,909	6,7954	79,22	6,115	0,259	0,0456	0,8466	0,1533

Source: prepared by the authors (2019).

The result of the T-test of the comparison of means is available in Table 9. Again, we consider the H0 hypotheses (there is no difference between company B's page posts on Facebook and Instagram regarding sentiment) and H1 (there is a difference between the posts on company B's Facebook and Instagram pages regarding sentiment).

Table 9 - Difference between polarity means on company B's pages

	Method Sentistrength		Method Umigon	
	Facebook	Instagram	Facebook	Instagram
Mean	0,350726	0,258425	0,238471	0,110473
Variance	0,290236	0,300498	0,29777	0,241573
Observations	5443	6617	5443	6617
P-value:	0,0000		0,0000	
Df	0		0	
t Stat	11686		11073	
P(T<=t) one-tail	9,288946		13,40207	
t Critical one-tail	9,16E-21		6,09E-41	
P(T<=t) two-tail	1,644984		1,644991	
t Critical two-tail	1,83E-20		1,22E-40	

Source: prepared by the authors (2019).

Since the P (T <= t) value is less than 0.05, and t Stat exceeds the critical t value, we reject H0. Therefore, we can affirm differences in the average sentiment polarity of the posts. However, the contradictory findings between the two companies make it difficult to conclude the presence of an apparent pattern in the tendencies of manifestations of emotions in the two different social media.

The apparent prevalence of positive comments over negative comments on the pages of the two companies on different social media may indicate a high level of consumer satisfaction with the provided service, as shown by the work of Miranda and Sassi (2014). It is important to note that the tendency of the



SentiStrength method to classify texts as a neutral class in a three-class analysis reported by Ribeiro *et al.* (2016) is not observed in this research when compared with the Umigon method - which received, on average, the best performance measures for this type of analysis. However, we can observe that, in practically all the analyzed databases, more comments were classified as neutral by the Umigon method.

4.2 Companies' responsiveness

Table 10 shows the general data on the number of companies' responses to customer comments. In both media, there is a low rate of response to comments. The large volume of comments on social media can increase the cost or even make it impossible to show attention to all or even most of the comments, despite the benefits in engagement and responsiveness brought by the firm-consumer dialogue pointed out by Gallagher and Ransbotham (2010) and the positive impacts on consumer satisfaction pointed out by Agnihotri *et al.* (2016).

Table 10 - General indices of companies' responsiveness to comments

	Company A		Company B	
	Instagram	Facebook	Instagram	Facebook
Total answered comments	1035	769	1262	1403
Total answered people	981	631	1249	1136
Percentage of answered comments	15,50%	13,65%	19,42%	26,15%

Source: prepared by the authors (2019).

The differences in the percentages of answered comments show that companies have given different attention to the comments of each media since company A responds more to Instagram comments (Tables 11 and 12). In contrast, company B does the opposite (Tables 13 and 14). The media that receive the most attention from companies are the ones where the highest positive comments are present, respectively. This fact may indicate a possible correlation between a company's responsiveness in online communication and the polarity of emotions expressed in comments related to the company, and consequently in consumer satisfaction (MIRANDA; SASSI, 2014), which is in agreement with the work of Agnihotri *et al.* (2016).

Table 11 - Comments and people answered by Company A (Instagram)

	Method	Positive	Negative	Neutral	P/N	P/Total	NG/Total	P/(P+NG)	NG/(P+NG)
Comments	SentiSTR	571	59	405	9,6779	0,55169	0,057	0,9063	0,09365
	Umigon	460	103	472	4,4660	0,44444	0,099	0,8170	0,18294
People	SentiSTR	558	58	365	9,6206	0,56881	0,059	0,9058	0,09415
	Umigon	451	101	429	4,4653	0,45974	0,103	0,8170	0,18297

Source: prepared by the authors (2019).

Table 12 - Comments and people answered by Company A (Facebook)

	Method	Positive	Negative	Neutral	P/N	P /total	NG /total	P/(P+NG)	NG/(P+NG)
Comments	SentiSTR	387	31	351	12,483	0,50325	0,0403	0,9258	0,07416
	Umigon	332	80	357	4,15	0,43173	0,1040	0,8058	0,19417
People	SentiSTR	354	20	257	17,7	0,56101	0,0316	0,9465	0,05347
	Umigon	309	58	264	5,3275	0,48969	0,0919	0,8419	0,15803

Source: prepared by the authors (2019).

We observe that the proportion of positive comments answered to the total answered is more significant than that of the total positive with the total sample. Thus, the Instagram page's administrators tend to interact with comments that express positive emotions. This trend is valid from the view of digital marketing, knowing the benefits of positive comments in promoting the company (EVANGELISTA; PADILHA, 2014). Thus, companies can engage more with people who comment positively to encourage them to continue speaking out about the company, promoting word-of-mouth marketing, as Gallagher and Ransbotham (2010) pointed out.

This discrepancy is apparent with greater intensity on Facebook posts, in which the proportion of total positive comments to total comments is 24.57%. The same proportion for the sample of comments answered is 50%, approximately twice the total ratio. Compared to the table of emotions' polarity, we observe that, even with a higher rate of responses, the moderators of company B's page show a slight tendency, although less intense, to respond more to comments classified as positive over negative ones.

The discrepancy between the proportions of positive to the total answered comments could also be observed less intensely in company B. Therefore, in addition to the advantages explained by Gallagher and Ransbotham (2010) and considering the findings of the research by Coelho and



Gosling (2015), it is possible to state that responding to positive comments is a way of maintaining a good relationship with users with a greater potential for engagement, a fact that can be identified in this research, when, in addition to the identified preference by positive comments, it is also apparent that the media that receive the most comments from the company are the ones that have the most positive comments from their customers.

Table 13 - Comments and people answered by Company B (Instagram)

	Method	Positive	Negative	Neutral	P/N	P /total	NG /total	P/(P+NG)	NG/(P+NG)
Comments	SentiSTR	519	46	697	11,282	0,41125	0,0364	0,9185	0,08141
	Umigon	313	84	865	3,7261	0,24801	0,0665	0,7884	0,21158
People	SentiSTR	354	20	257	17,7	0,56101	0,0316	0,9465	0,05347
	Umigon	309	58	264	5,3275	0,48969	0,0919	0,8419	0,15803

Source: prepared by the authors (2019).

Table 14 - Comments and people answered by company B (Facebook)

	Method	Positive	Negative	Neutral	P/N	P /total	NG /total	P/(P+NG)	NG/(P+NG)
Comments	SentiSTR	754	29	620	26	0,53742	0,0206	0,9629	0,03703
	Umigon	720	87	596	8,2758	0,51318	0,0620	0,8921	0,10780
People	SentiSTR	354	20	257	17,7	0,56101	0,0316	0,9465	0,05347
	Umigon	309	58	264	5,3275	0,48969	0,0919	0,8419	0,15803

Source: prepared by the authors (2019).

In that way, companies engage customers who tend to comment positively on their social media pages. However, Gallagher and Ransbotham (2010) and Evangelista and Padilha (2014) also point to the need to monitor and respond to negative comments, given their high power to damage the company's image.

4.3 Aspects of improvement

The results indicate that, when manifesting some dissatisfaction on the page of company A, users sought to associate the company figure with various types of negative adjectives, as shown in Table 15. Words such as "service," "problem," "company," and "client," despite having expansive meanings in this context, present a north for further analysis since the high incidence of these

words and their associated adjectives indicate that a considerable part of negative comments is related to problems in the company's services, especially in customer service, prompting the question of which aspects of that service would be presenting problems. The presence of negative comments regarding the company represents aspects of consumer-consumer or consumer-to-business communication (GALLAUGHER; RANSBOTHAM, 2010) that can negatively impact consumer confidence in the company, an element considered by Forouzandeh, Soltanpanah and Sheikahmadi (2014) as the biggest advantage of content marketing efforts.

Table 15 - Associated nouns and adjectives in the comments on company A's pages

Word	Incidence	Associated Adjectives
company	195	terrible, serious, worst, horrible, crap, garbage, biggest, reliable, crook, aerial, successful, corrupt, cancel, irresponsible, dignified, honest
ticket	162	expensive, cheap, problematic, integral, failed, area, super, aerial
problem	134	@user ¹ , very serious, serious, impossible, unique, boring, ridiculous
Site	104	parallel, difficult, loud, pitiful
trip	104	next, aerial, whole, good, direct, first
tickets	95	aerial, guys, good
money	87	@user ²
purchase	82	Final, urgent
service	81	terrible, sale, terrible, horrible, decent, easy, bad, different, worst
package	80	cheap, sinister
contact	73	with me, difficult, urgent
customer	48	neglect (with), direct, available
reservation	46	urgent
telephone	42	App
flight	42	Canceled

Source: prepared by the authors (2019).

Words like "ticket," "website," "trip," "money," and "package," among others and their associated adjectives, indicate more specific service aspects that have problems – for example, possible problems with the ticket. Also, the ticket price is a commonly mentioned element. Through the individual analysis of the comments, it was possible to identify the recurrence of the situation reported by

¹ User name censored for privacy.

² User name censored for privacy.



Coelho and Gosling (2015) of the presence, to a lesser extent, of travelers who engage in reporting complaints and negative occurrences in their experience.

The content of 100 reviews of the company's purchase and after-sales was analyzed. Of this sample, 29 people presented problems regarding the ticket not being sent by e-mail until the moment of the comment or problems in the ticket issuance on the website. It is important to note that most of these cases are related to fraud, reported by company A and customers. The fraudster impersonates the company, making a fake sale through the Facebook chat system. Thus, reports of what happened could be observed primarily in Facebook comments. The company made it clear when responding to comments that they do not make sales through Facebook - a necessary movement, as explained by Gallagher and Ransbotham (2010), in correcting mistakes and mitigating damages. However, the analysis reveals misunderstandings in which people associated the company with fraud, potentially damaging the company's image (EVANGELISTA; PADILHA, 2014). In addition, 21 people in the sample had problems where the airline company canceled the ticket or the route or date was changed, causing various inconveniences to customers.

It is important to point out that in some cases, customers reported a relation to the Avianca airline's recent financial and operational problems (BRANCO; CAVALCANTI; DOCA, 2018) and the company's difficulties in accomplishing the rerouting of flights in order to satisfy its entire clientele. Thirteen people reported not getting a refund, which, in most of the reports, was related to changes in the ticket's date carried out by the responsible company. Eleven people expressed their opinion that the ticket was expensive, eight people had difficulties purchasing the ticket, either on the website or in the app, and the same number had difficulties canceling or rescheduling. Four people showed dissatisfaction with the fee charged for rescheduling the trip. We can also observe in a smaller quantity: problems with undue collection, tickets without the seat's number, and other problems in the after-sales. Thus, Moghaddam and Esther (2010) reported that natural language processing could help extract opportunities for improvement in those services. Table 16 exemplifies comments that report the observed problems.

Through the analysis of 50 incidences of the word "site" in negative comments, it is possible to observe that a large part of the comments also

contained the word "ticket" in order to indicate that previously presented problems, such as difficulties in the purchase, cancellation or rescheduling, possibly have some relation to the functioning of the website, since customers purchase the tickets and packages through the company's website or official app. Additionally, reports of instability on the site or blocked customer access are present with a very low incidence.

The analysis of negative comments containing the words "trip," "money," or "package" (50 recurrences of each word) also indicates, in general, the presence of the same problems identified in the previous analysis, which states that in the comments containing the word "trip," there is a greater incidence in reports of problems such as changes in the flight and the trip's route or date. In addition, comments with the word "money" highlight problems with cancellations and refunds.

Table 16 - Examples of ticket problems reported by company A customers

Problem	Example of comment extracted from the page
ticket not sent	please I bought 3 tickets to João Pessoa by bank slip, paid on 12/03 and I still have had no answer so I'm desperate help me please, I can't buy another ticket ??
	I was blocked on Messenger where I handled tickets, for questioning other fees they said they forgot to launch.
difficulty of canceling	(Company reply) User ³ , we do not make sales through Facebook Messenger or any other social media. Only through the website, app or telesales at the number available on our website.
	I had a death in the family and you didn't cancel the ticket
cancelled ticket	Dear company A, I came to inform you that I have MY WEDDING in the city of Fortaleza on April 28th, and many of my guests bought tickets from your company on the GRU - FOR route to attend the event. INCLUDING MY PARENTS !! The flights were canceled and you did not guarantee relocation. (...)
lack of refund	Too bad I made a purchase, canceled within 24 hours and so far I have not received a refund! Shame on you @company A owing me a refund of R\$ 6 thousand over 1 month !!!!!
expensive ticket	I bought a ticket from Acre to Santa Catarina, I paid R \$ 851.03 I think the price is very high, I have friends who bought from other companies much more cheaply, next time I will do more research to avoid falling into this "trap".
	I can't buy tickets on the website.

³ User name is censored for privacy.



difficulty on buying tickets	Company A's website has been redone to confuse the customer. Simply horrible!!!!
rescheduling fee	Watch out people! Company with terrible service We changed the trip's date, it was imposed a charge of almost R \$ 500.00 of fines and fees and when the e-mail with the exchange confirmation came, the date was wrong We opened a complaint at the company and they said that the conversation indicated that the mistake was ours (...)

Source: prepared by the authors (2019).

It is interesting to point out that from this sample, 19 comments are from people requesting the resolution of a specific case in which a woman who is a pastor and singer did not obtain her refund, which exemplifies the power of individuals who stand out due to their influence in certain groups to move people to the manifestation of causes, even though punctual ones, and impact companies' digital image, both negatively and positively, following the research by Silva and Tessarolo (2016).

Analogously to the case of company A, there is a high incidence of nouns such as "company," "problem," and "service" in company B's negative comments (Table 17), which indicates possible problems in the services provided by it. In addition, the high incidence of words such as "trip," "package," "ticket," "agency," "ship," "cruise," and "hotel," among others, guide the search for a content analysis of comments containing these words.

Table 17 - Associated nouns and adjectives in comments on company B's pages

Word	Incidence	Associated adjectives
trip	76	first, next, good, aerial, great, unforgettable, special
package	53	expensive, complete, neglect
dream	44	old, cool, silly, big
company	41	aerial, Avianca, garbage, best, worst, single, partner, show, dissatisfied
problem	37	urgent
passage	30	aerial, expensive, complete
place	24	full, good, expensive, spectacular
service	23	best, terrible, bad
saudades ⁴	23	guide, girls, unforgettable

⁴ There is no corresponding English word for the portuguese word "saudades", which means the longing for something that is in the past.

agency	21	worst terrible, multi-brands, place, best
fear	20	now
ship	15	sovereign, good, old
packages	14	new
daughter	14	
cruise	14	cheap, awesome
money	14	lack, lost
company	13	lack
city	12	small, worst, big
hotel	12	luxurious
valor	11	final, high

Source: prepared by the authors (2019).

An analysis of 50 negative comments containing the word "trip" indicates that, in general, part of the problems presented above are identified for this company. 16 of the comments indicated a problem in which the company canceled the trip or flight and rescheduled, causing complications for customers, including problems in reversing the amount paid. In most of those cases, a relation between the issues and the company Avianca was reported, a fact already mentioned previously.

It is important to note that seven people desire to travel using the company's services but cannot due to financial conditions or other reasons. From this sample, there are three comments in which customers report their intention to hire the company's services, but "fear" manifested after the negative comments. These reports exemplify the results of the study "The 2012 traveler" (THINK..., 2012), which states that, when planning their trips, certain people seek criticism and opinions in online comments, in addition to representing signs of a possible loss of confidence in the company's services. In addition, other punctual complaints are present regarding the price of the trip and the fees for rescheduling the date.

The analysis of 50 incidences of the word "package" exposed the problem with the company Avianca, representing a large part of the complaints, and identified four complaints concerning the lack of package options, specifically for the following destinations: the city of Fortaleza, the continent of Africa, the Brazilian state of the Northeast and from Ilhéus to Itacaré. Monitoring and identifying this type of comment is in line with the research of Zeng and Gerritsen (2014) and Fotis, Buhalis, and Rossides (2012), which reinforces the



importance of social media in the search for information when making decisions. For example, comments expressing a demand for experiences in specific locations can support decisions on preparing travel packages. The analysis of 30 incidences of the word "ticket," 21 of the word "agency," and 15 of the word "ship" also pointed out some additional, although less reported, problems, such as after-sales issues, unsatisfactory service and difficulties in getting the chargeback for canceled flights. Examples are available in Table 18.

Table 18 - Examples of problems related by company B's customers

Problem	Example of comment extracted from the page
canceled or rescheduled flights	@user ⁵ My trip was canceled the day before departure 04/27. I was supposed to embark today 04/28. I don't want to discourage you, but if they are negligent with you like they were with me, your trip will be canceled.
	I've been scheduling the vacation trip for 3 months, the day before departure you cancel my trip and just ask to go to the agency for a refund. It is a total disregard for the client who trusted you. Now all there is to do is unpacking, sitting and crying, because my dream of visiting another place has been canceled. Thank you for making my long-awaited vacation the worst of all.
	Company B my flight was canceled, my trip is 05/05, what do I do?
desire to travel	A dream, an unforgettable trip, only if sponsored by company B, because I have no conditions 😞
fear for negative comments	Mercy, I read so many negative comments that I will no longer be choosing company B to provide my trip.
problems with Avianca	@user ⁶ did you get on board? I bought 4 packages with company B for June 30th and the flights are from Avianca. I have already contacted them, as I don't want to arrive at the airport and just go back, specially with 2 small children
problems on after sales service	Bad after sales service ... NEVER AGAIN.
No chargeback	Solve my problem and refund my money

Source: prepared by the authors (2019).

The comments containing the words "dream" and "saudade" indicate the desire of users to know the places and experiences announced by the company,

⁵ User name censored for privacy.

⁶ User name censored for privacy.

as well as the feeling of longing for past experiences. These comments demonstrate successful situations in companies' content marketing efforts through posts, arousing consumers' desire for possible products and services offered by the company (FOROUZANDEH; SOLTANPANA; SHEIKHAHMADI, 2014). It is interesting to note that there is a possible tendency in the sentiment analysis algorithms to classify comments that express longing and desire as being negative since there is a subjectivity of the meaning of these words concerning the different contexts of a sentence.

Analyzing Table 19, we can observe that with the use of Watson's automated keyword extraction tool, it is possible to identify the same relevant words concerning the company's service aspects, identified with the automated extraction of nouns and adjectives.

Table 19 - Keywords with the highest incidence in comments (using Watson NLU)

Company A	Incidence	Company B	Incidence
Money	200	company B	59
User ⁷	180	Company B(id)	28
Ticket	118	package	26
Company	106	dream	23
Problem	78	to	15
Contact	77	gente ⁸	15
You	76	problem	15
Tickets	76	ticket	13
Purchase	65	company	13
Package	62	problems	13
Site	60	desire	12
Client	47	money	11
Telefone	44	people	11
No	44	promotion	10

Source: prepared by the authors (2019).

4.4 Watson NLU's Performance on Emotion Classification Test

Finally, using the IBM Watson module to analyze emotions, we present the classification performance test results for three classes of emotions (Table 20). The

⁷ User name censored for privacy.

⁸ The portuguese word "gente" could be translated to English both as "people" or as "us," depending on the context.



database contains a balanced and an unbalanced corpus. We can observe higher accuracy than the results described by Aguiar *et al.* (2018) using the same database, which may indicate the development of the tool's classification algorithms over the years, expanding the relevance of the method used in the present research.

Table 20 - Performance metrics on Watson NLU emotion ranking

Corpus	Total Successes	Errors	Accuracy	Precision	Recall	F1-Measure
unbalanced	1257	762	0.622585438336	0,655864	0,776	0,70693
balanced	317	181	0.636546184739	0,545664	0,702992	0,565697

Source: prepared by the authors (2019).

5. Final considerations

Given the results presented, it is possible to state that the research achieves its proposed objectives. The analysis of the polarity of emotions identified an apparent prevalence of positive comments on the pages of the two companies, which may indicate a good level of consumer satisfaction. In addition, the results of the T-Student test indicate that there may be differences in the mean polarity of feeling when comparing the two social media while analyzing each company individually. However, the contradictory results observed while analyzing the two companies together counteract the hypothetical existence of a clear trend, specifically from the users of one of the two social networks, in the polarity of the feelings expressed.

There was also a positive correlation between the companies' online responsiveness and the polarity of feelings expressed in customer comments. However, there is a tendency for the administrators of the analyzed companies' pages to interact more with comments that express positive feelings, which may act to the detriment of a company's levels of responsiveness, reducing the positive impacts on consumer satisfaction.

The analysis of the nouns, adjectives, and other keywords extracted from the comments indicates the presence of aspects of the provided service that could be improved, particularly the ability of companies to manage unforeseen events caused by cancellations, changes in flights, and other problems of partner airlines, in order to ensure the satisfaction of its customers. In one of the

companies, there is also a problem with fraud carried out by third parties using the company's name illegally, negatively influencing its operations and image.

The present work offers relevant contributions for researchers and managers, incorporating natural language processing techniques in extracting content generated by users on social media and identifying signs of problems and possible points for improvement in offered services and products. In addition, the identified trends in consumer behavior on social media can support decision-making concerning digital marketing strategies, including monitoring customer comments, while providing insights for future research on consumer behavior on social networks.

Regarding the tourism sector, we recommend that the agencies invest in better ways to manage unforeseen events caused by flight cancellations and other problems of partner airlines since this study identified great dissatisfaction about this.

Agencies can also benefit by increasing their responsiveness by instructing their social media moderators to pay greater attention to their clients' negative and neutral comments.

It is important to note that there are limitations in research inherent mainly to the technology used. First, the great difficulty in obtaining the approval of the companies that own the social media to use APIs for automated data collection made manual data collection a slow process. It limited the number of comments collected over the time available for conducting the research. Second, We can observe that there is currently much room for improvement in sentiment analysis algorithms and other natural language processing tools available for the analysis of texts in Portuguese, a fact possibly related to little effort in the creation of Portuguese text datasets labeled for the training of machine learning models and the development of updated lexical dictionaries of the language.

Given the above, we emphasize the importance of future research on the development and improvement of natural language processing tools in the Portuguese language since, as previously reported, there is an expressive adoption of social media in Brazilian daily life, which, in turn, represents a massive generation of data that can be used as research sources, increasing the positive impact not only in the productive business sectors but also in society as a whole. With concern to social media applied in the tourism sector, future research may

analyze aspects such as the correlation of the polarity rates of emotions expressed in comments with the different constructs of quality and consumer satisfaction present in the literature in order to enrich the theoretical bases for the application of more studies using sentiment analysis techniques in this environment. Also, more studies comparing the two social networks when it comes to the polarity expressed in their comments may provide us with a better understanding of the average user of the two networks, as well as insights into how different characteristics in the style of posting on a specific social network may influence the polarity of the response comments, in order to assist companies' decisions when it comes to online communication.

References

- AGNIHOTRI, R. *et al.* Social media: Influencing customer satisfaction in B2B sales. **Industrial Marketing Management**, [s.l.], v. 53, [s.n.], p. 172–180, 2016.
- Aguiar, E. *et al.* Análise de Sentimento em Redes Sociais para a Língua Portuguesa Utilizando Algoritmos de Classificação. *In.*: Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, 36., 2018, Porto Alegre. **Proceedings...** Porto Alegre: [s.n.], 2018. p. 393-406,
- AMARAL, F. *et al.* Comentários no TripAdvisor: do que falam os turistas?. **Dos Algarves**: a multidisciplinary e-journal, [s.l.], v. 2, n. 26, p. 47-67, 2017.
- ARAUJO, M. *et al.* An evaluation of machine translation for multilingual sentence-level sentiment analysis. *In.*: PROCEEDINGS OF THE 31ST ANNUAL ACM SYMPOSIUM ON APPLIED COMPUTING, 31., 2016, Pisa. **Proceedings...** Pisa: ACM, 2016. p. 1140–1145.
- ARAUJO, M. *et al.* iFeel: a system that compares and combines sentiment analysis methods. *In.*: THE 23RD INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 23., 2014, Seoul. **Proceedings...** Seoul: ACM Press, 2014. p. 75-78.
- ARAUJO, M.; GONÇALVES, P.; BENEVENUTO, F. Measuring sentiments in online social networks. *In.*: PROCEEDINGS OF THE 19TH BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 19., 2013, Salvador. **Proceedings...** Salvador: ACM Press, 2013. p. 97-104.
- ARAUJO, R. Marketing científico digital e métricas alternativas para periódicos: da visibilidade ao engajamento. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 20, n. 3, p. 67-84, 2015.
- BALTES, L. Content marketing-the fundamental tool of digital marketing. **Bulletin of the transilvania university of brasov**. Series V: economic sciences, Romania, v. 8, n. 2, p. 111-118, 2015.

BARDIN, L. **Análise de conteúdo**. Lisboa: Edições, 2011. v. 70. (Edição revista e actualizada)

BRADBURY, K. The growing role of social media in tourism marketing. *In.*: BRADBURY, K. **Blog bur y**. [s.l.]: [s.n.], 2011. Available in: <https://kelseybradbury.weebly.com/uploads/1/0/9/2/10927387/tourismsocialmedia-comm427.pdf>. Access in: 07 jun. 2019.

BRANCO, L.; CAVALCANTI, G.; DOCA, G. Avianca Brasil pede recuperação judicial por risco de paralisar suas operações. **O Globo**, Rio de Janeiro, [s.n.], [s.n.], 17 abr. 2019. Economia Available in: <https://oglobo.globo.com/economia/avianca-brasil-pede-recuperacao-judicial-por-risco-de-paralisar-suas-operacoes-23297762>. Access in: 07 jun. 2019.

CHOI, J.; TETREULT, J.; STENT, A. It depends: dependency parser comparison using a web-based evaluation tool. *In.*: PROCEEDINGS OF THE 53RD ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 58.; THE 7TH INTERNATIONAL JOINT CONFERENCE ON NATURAL LANGUAGE PROCESSING, 7., 2015, Beijing. **Proceedings...** Beijing: ACL Anthology, 2015. p. 387-396. (Volume 1: Long Papers).

COELHO, M. F.; GOSLING, M. S. **Comentar bem ou mal na internet?** O engajamento de viajantes em reviews de hotéis. Revista: Turydes Revista Turismo y Desarrollo 2015.

DWIVEDI, M.; SHIBU, T.; VENKATESH, U. Social software practices on the Internet: Implications for the hotel industry. **International journal of contemporary hospitality management**, [s.l.], v. 19, n. 5, p. 415–426, 2007.

EVANGELISTA, T.; PADILHA, T. Monitoramento de postagens sobre empresas de e-commerce em redes sociais utilizando análise de sentimentos. *In.*: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING (BraSNAM), 2., 2014, Porto Alegre. **Proceedings...** Porto Alegre: SBCOPENLIB, 2014.

FOROUZANDEH, S.; SOLTANPANAH, H.; SHEIKHAHMADI, A. Content marketing through data mining on Facebook social network. **Webology**, [s.l.], [s.n.], n. 1, p. 1-11, 2014.

FOTIS, J.; BUHALIS, D.; ROSSIDES, N. Social media use and impact during the holiday travel planning process. *In.*: Fuchs, M.; Ricci, F.; Cantoni, L. **Information and Communication Technologies in Tourism**. Sweden: Springer, 2012. Available in: https://link.springer.com/chapter/10.1007/978-3-7091-1142-0_2. Access in: 15 mar. 2019.

GALLAUGHER, J.; RANSBOTHAM, S. Social media and customer dialog management at starbucks. **MIS Quarterly Executive**, [s.l.], v. 9, n. 4, 2010.

GONÇALVES, P. *et al.* **Comparing and combining sentiment analysis methods**. *In.*: PROCEEDINGS OF THE FIRST ACM CONFERENCE ON ONLINE SOCIAL NETWORKS, 1., 2013, [s.l.]. **Proceedings...** [s.l.]: ACM Press, 2013.

GONZALES, M.; LIMA, V. Recuperação de informação e processamento da linguagem natural. *In.*: Congresso da sociedade Brasileira de computação, 23., 2003, [s.l.]. **Proceedings...** [s.l.]: [s.n.], 2003. p. 347-395

HEARST, M. **Untangling Text Data Mining**. *In.*: PROCEEDINGS OF THE 37TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS ON COMPUTATIONAL LINGUISTICS, 37., 1999, College Park. **Proceedings...** College Park: University of Maryland, 1999. p. 3-10.

LEUNG, D. *et al.* Social media in tourism and hospitality: a literature review. **Journal of Travel & Tourism Marketing**, [s.l.], v. 30, n. 1-2, p. 3-22, jan. 2013.

LIMA, V. M. **Engajamento do consumidor em uma comunidade virtual de marca**. 2014. 103 f. Dissertação (Mestrado em Gestão Empresarial) – Escola Brasileira de Administração Pública e de Empresas, Fundação Getulio Vargas, Rio de Janeiro, 2014.

LUCAS, A. S. **Top 10 maiores agencias de viagens do Brasil**. [s.l.]: Top10+, c2013. (Luxo). Available in: <https://top10mais.org/top-10-maiores-agencias-de-viagens-brasil/>. Access in: 15 mar. 2019.

MALHOTRA, N. **Pesquisa de marketing: uma orientação aplicada**. 3. ed. Porto Alegre: Bookman, 2001.

MIRANDA, M. D.; SASSI, R. J. Using sentiment analysis to assess customer satisfaction in an online job search company. **Business Information Systems Workshops**. [s.l.], v. 183, [s.n.], p. 17-27, 2014.

MOGHADDAM, S.; ESTER, M. Opinion digger: an unsupervised opinion miner from unstructured product reviews. *In.*: PROCEEDINGS OF THE 19TH ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 19., 2010, Toronto. **Proceedings...** Toronto: ACM Digital Library, 2010, p. 1825-1828.

MORENO, A. C. **Análise de sentimentos na classificação de comentários online aplicando técnicas de text mining**. 2015, 72 f. Dissertação (Mestrado em Sistemas Integrados de Apoio à Decisão) – Departamento de Ciências e Tecnologias de Informação, Instituto Universitário de Lisboa, Lisboa, 2015.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. **Foundations and Trends in Information Retrieval**, [s.l.], v. 2, n. 1-2, p. 1-135, 2008.

PANTELIDIS, I. Electronic meal experience: a content analysis of online restaurant comments. **Cornell Hospitality Quarterly**, [s.l.], v. 51, n. 4, p. 483-491, ago. 2010.

REIS, J. *et al.* Uma abordagem multilingue para análise de sentimentos. *In.*: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING (BraSNAM), 4., 2015, Porto Alegre. **Proceedings...** Porto Alegre: SBCOPENLIB, 2015.

- RIBEIRO, F. N. *et al.* Sentibench - a benchmark comparison of state-of-the-practice sentiment analysis methods. **EPJ Data Science**, [s.l.], v. 5, n. 23, 2016.
- SANTOS, G. C. O. *et al.* As redes sociais e o turismo: uma análise do compartilhamento no Instagram do Festival Cultura e Gastronomia de Tiradentes. **Revista Iberoamericana de Turismo (RITUR)**, Penedo, v. 7, n. 2, p. 60–85, 2017
- SCHULTZ, B. B. Levene's test for relative variation. **Systematic Zoology**, [s.l.], v. 34, n. 4, p. 449–456, dec. 1985.
- SHARDA, R.; DELEN, D.; TURBAN, E. **Business intelligence: a managerial perspective on analytics**. Hoboken: Prentice Hall Press, 2013.
- SILVA, C.; TESSAROLO, F. **Influenciadores digitais e as redes sociais enquanto plataformas de mídia**. In.: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 39., 2016, São Paulo. **Proceedings...** São Paulo: INTERCOM, 2016.
- SOUSA, A. R. Processamento automático de línguas naturais: Um estudo sobre a localização do IBM Watson™ para o português do Brasil. 2017. 76 f. Monografia (Bacharel em Línguas Estrangeiras Aplicadas ao Multilinguismo e à Sociedade da Informação) – Instituto de Letras, Universidade de Brasília, Brasília (DF), 2017.
- STICH, V.; EMONTS-HOLLEY, R.; SENDEREK, R. Social media analytics in customer service: a literature overview - an overview of literature and metrics regarding social media analysis in customer service. In.: 11TH INTERNATIONAL CONFERENCE ON WEB INFORMATION SYSTEMS AND TECHNOLOGIES, 11., 2015, [s.l.], **Proceedings...** [s.l.]: SCITEPRESS - Science and Technology Publications, 2015.
- THEVENOT, G. Blogging as a social media. **Tourism and hospitality research**, [s.l.], v. 7, n. 3-4, p. 287–289, 2007.
- THINK with Google. **The 2012 Traveler**. USA: Google: Ipsos MediaCT, 2012. Available in: https://www.thinkwithgoogle.com/_gs/documents/682/the-2012-traveler-research-studies.pdf. Access in: 11 ago. 2019.
- VICENTE, M. Uma imagem vale mais que mil palavras? Humans Of New York e a febre de páginas que contam histórias de anônimos através de fotografias no Facebook. **Revista Mangaio Acadêmico**, [s.l.], v. 1, n. 1, p. 01-11, 2016.
- YE, Q.; LAW, R.; GU, B. The Impact of Online User Reviews on Hotel Room Sales. **International Journal of Hospitality Management**, [s.l.], v. 28, n. 1, p. 180–182, mar. 2009.
- ZENG, B.; GERRITSEN, R. What do we know about social media in tourism? A review. **Tourism Management Perspectives**, [s.l.], v. 10, [s.n.], p. 27–36, Apr. 2014.

APPENDIX

Script operation in Facebook spreadsheets: (1) The rows whose following lines contain values in the "answer" column are searched. (2) Once the script finds a line, the number of answers in the comment is counted until it reaches the following line that does not contain the cited column. (3) The search for comments from the company's official user is carried out on these lines. If it finds multiple recurrences of company comments, it extracts from each company comment the list of entities in the text. (4) Each entity in this list is searched in the answers users column to search for the answered user. (5) These data are saved in a row of another spreadsheet containing: the comment, the polarity of emotions (according to the Sentistrength and Umigon methods), and the company's response. It repeats the procedure on the following line, non-classified as an answer.

Script operation in Instagram spreadsheets: (1) Comments in a post are counted. (2) A search is performed for each line in company comments on a post. (3) Once it identifies the company's comment, the words that start with the character '@' are searched in the text since it contains the answered user's name. (4) Each word found is searched in the post's user list. (5) If it finds a user, these data are saved in a row of another spreadsheet containing: the comment, the polarity of emotions (according to the Sentistrength and Umigon methods), and the company's response. It repeats the procedure on the following line, not classified as an answer.