

## Introduction to the Analysis of Tourism Services through Feelings Mining

Hector F. Gomez <sup>1\*</sup>, Edwin Fabricio Lozada <sup>2</sup>, Carlos Eduardo Martínez <sup>2</sup>, Freddy Patricio Baño <sup>2</sup>,  
Gustavo Adolfo Álvarez <sup>2</sup>, Walter V. Culque <sup>2</sup>, Natalia Soledad Bustamante-Sánchez <sup>3</sup>,  
Rosario Estefanía Sánchez Cevallos <sup>3</sup>

<sup>1</sup> *Facultad de Ciencias Humanas y de la Educacion, UHuachi- Ambato, ECUADOR*

<sup>2</sup> *Carrera de Sistemas, Universidad Regional Autónoma de los Andes–Uniandes-Km. 5½ vía a Baños, ECUADOR*

<sup>3</sup> *Universidad Técnica Particular de Loja, Departamento de ciencias Administrativas, Sección de Hotelería y Turismo, ECUADOR*

\*Corresponding Author: [hf.gomez@uta.edu.ec](mailto:hf.gomez@uta.edu.ec)

**Citation:** Gomes, H. F., Lozada, E. F., Martínez, C. E., Baño, F. P., Álvarez, G. A., Culque, W. V., Bustamante-Sánchez, N. S. and Cevallos, R. E. S. (2018). Introduction to the Analysis of Tourism Services through Feelings Mining. *Journal of Information Systems Engineering & Management*, 3(4), 33. <https://doi.org/10.20897/jisem/3944>

**Published:** November 10, 2018

### ABSTRACT

One of the sources of information to detect the level of customer satisfaction is the set of reviews they leave on websites such as Trip Advisor. A problem arises when there is a large number of reviews and administrators have difficulty identifying those that reflect the client's feelings from the reviews. For this study, some applications generated in the Python programming language were used in order to assign an excellent and bad criterion by adding a feeling textually. The Sentistrength classifier was also used, which analyzes text to return a value of positive or negative sentiment; these two contexts were used mathematically and statistically to obtain referential data based on an analysis of the ROC curve with values of true positives and true negatives as well as false positives and false negatives.

**Keywords:** tourism, semantic, orientation, opinion

### INTRODUCTION

The tourism industry considers several sources of information to improve its offer to tourists, and needs to measure the level of satisfaction of its customers when using different hotel services. This study considers that it is possible to determine the level of satisfaction of hotel service clients through their reviews, identifying the polarity of text phrases in these reviews. The objective is to classify the reviews automatically. This could make it possible to suggest improvements in hotel services in relation to the issue addressed therein. Several research studies have been carried out to try to detect the polarity of a given text.

“This research makes reference to the study of human emotions in front of the satisfaction of service received in a hotel, it is also sought to compare the results obtained between two binary classifiers PMI (Point Mutual Information) and Sentistrength. PMI is a method used in information theory and statistics to measure the correlation between two variables” (Aurchana et al., 2014). “Sentistrength is a lexicon-based classifier, and also uses lexicon-independent linguistic information, as well as rules for detecting the strength of sentiment in informal short text written in English” (Choo et al., 2015).

Emotions are the product of feelings provoked at a certain moment according to the level of satisfaction perceived. These emotions can include: joy, sadness, displeasure, fear, etc. These emotions can be assigned a grammatical structure that can be interpreted numerically.

- (1) <http://sentistrength.wlv.ac.uk/results.php?text=I+really+love+you+but+dislike+your+cold+sister.&submit=Detect+Sentiment>

The results were subjected to ROC analysis, and it was observed that the Polarity-Emotion classifier allowed better than random classification results to be obtained, so we concluded in the present paper that this classifier performed well, while the PMI and Sentistrength classifiers performed poorly, worse than randomly. In the end we concluded that other binary classifiers should be tested in order to identify the polarity in the text of hotel service reviews.

“In this research work we propose a methodology based on Point Mutual Information (PMI) to obtain the semantic orientation of the phrases that make up an opinion. Phrase selection is carried out according to the part of speech patterns of unsupervised classification proposed by” (Abirami et al., 2015). In this work, the semantic orientation is based on the orientation of words and phrases that can be consulted through Google's API. Phrases and their SO are used to train support vector machine (SVM), an algorithm that will be used to classify new opinions reflected by customers for a new or improved tourist package.

“In social networks, tourist opinions generally appear in blogs. The analysis of text in blogs may aid the classification of the polarity of an opinion, and by so doing help us to identify the real feedback of a traveller. However, blogs can often be very long. For this reason, we used micro blogs, which are smaller extracts of text. When analyzed, these may provide better results in the classification of polarity than when using the entire text. In the same way, for the classification we used SVM, thereby improving the calculation of precision. The latter was based on the calculation of confidence” (Schickert et al., 2015). In this work, we use these authors' proposals, seeing that the  $f$  was selected according to patterns. Therefore, not all the phrases were analyzed. For our training grid, we used a micro-blog. “The selection of phrases in this sense produced excellent results for the classification of the polarity of texts. For example the elimination of stop words helped improve classification criteria in the analysis of tweets” (Namahoot et al., 2015). “The text used in tweets, however, was organized with SVM. They likewise produced excellent results when polarity was found within the range of classification” (Prasath et al., 2015).

## ART STATE

“The information that flows during the execution process of a TP is very important for tourist operators who need to improve their services. Consumer behavior in social networks was examined using text mining techniques, which helped to classify it according to customer preferences and to obtain information about client segments that preferred one TP over another” (Aurchana et al., 2014). “With regard to customer opinions, we can also distinguish those that are spam” (Choo et al., 2015). Moreover, when they are processed as subjective opinions, they constitute noise in text mining process. Once the user's opinions have been validated, one can process the subjectivity and obtain classifications of tourist destinations (good, bad, average). “The VIKOR model is used to classify tourist destinations according to user opinions, to filter irrelevant commentaries, to extract feelings or sentiments, and to quantify them” (Abirami et al., 2015). We used the VIKOR principle, but with Semantic Orientation (SO) values close to 1, since it was necessary for the opinions to be close to semantic management. The multi-dimensionality of the model both enables the classification of places to be precise and aids the decision making process of the tourist operator and the consumer. The process of identifying valid opinions is very important in this study as it contributes to the semantic treatment of the data- making them as precise as possible for tourism operator consultancy. The customer opinions are not only used for text mining and for classifying a sentiment, but also for obtaining preferences. “When we examine social networks in relation to tourism and the evaluation of hotels, we aim to look at customer satisfaction and online management of the topics that are addressed in conversations with the clients” (Schickert et al., 2015). This is an area that the clients consider to be of vital importance since it links the conversation and client opinion with the entity. For example, the extract the hotel was clean becomes a positive opinion regarding the entity 'Hotel' by establishing a semantic textual relationship.

“These semantic representations have already been used for decision making in tourism. Combining semantics and information technology algorithms has thus enabled the classification of tourism documents” (Namahoot et al., 2015; Prasath et al., 2015). The design of semantic structures for tourism is related to consumer behavior, which is generally obtained by means of surveys that include measurements such as percentages, averages and standard deviations. These are also key aspects in tourism research. In this case, we worked with text mining techniques for semantic treatment and proposed the usage of the fuzzy logic since the client opinions were between (0 or 1) or had intermediate values which were expressed in the text with phrases such as more or less and average - a definition, which in our proposal, is used to describe the neutral state of a client with respect to an opinion

about an entity. The PANAS proposals and the levels of proposed pertinence in (Bakhtiyari and Husain, 2014) are utilized in our research with the aim of identifying an emotion and its level of pertinence.

## METHODOLOGY

The source of information for analyzing the polarity of emotions in the text were hotel services reviews in Trip Advisor, in the format shown below as an example (Gomez et al., 2018):

```
<Overall Rating>3.5
<Avg. Price>$172
<URL>http://www.tripadvisor.com/ShowUserReviews-g60878-d72586-r23256277-
Best_Western_Executive_Inn-Seattle_Washington.html
<Author>ardarvin
<Content>Deceptive Staff Deceptive front desk staff, claiming you cannot park on the street
between 10pm and 4am ... in order to try and get you to pay the $15 parking fee with them.
This is completely not true. Parking is free in Seattle on the street between 6pm and 8am.They
even put deceptive signage on the outside of the building facing the street, saying no parking
between 10pm and 4am... but this is wrong. It is a public street, and falls under the same Seattle
bylaws. Locals seem to know this and park there anyway to go to a nearby late-night club.
Anyway, save your money and park on the street (plenty of free space in this area), or stay at
the Travel Lodge next door which offers free parking.As for room...the organic fresh odor
neutralizer stuff they use is nauseously overpowering. Any attempts by them to be eco-friendly
were lost on us, as we had to blast the heat with the window wide open in order to try and
erate the room. Other than that, only minor complaints. It sucks that the room key is one big
advertisement for some pizza company, and that there is advertising and overbearing signage
in the room, but I guess this is to be expected for a major chain.
<Date>Jan 4, 2009

<No. Reader>-1
<No. Helpful>-1
<Overall>2
<Value>2
<Rooms>2
<Location>2
<Cleanliness>3
<Check in / front desk>1
<Service>2
<Business service>4
```

These comments describe the general opinion of clients regarding the service of a hotel, where it can be inferred that there are key words that denote the emotions of the tourist, even though different emotions can be shown in the tourist reviews. The present work is based on statistical tools that allow us to classify polarity based on key words or phrases written by tourists.

We compare the data processed in two statistical tools to evaluate their performance. The first is Pointwise Mutual Information (PMI), which allows text classification based on positive or negative polarity; this method takes a text phrase and compares it with databases available on the Internet to determine a review approach that attempts to classify the polarity of the text. The next method, Sentistrength, assesses whether a text has positive, negative or neutral sentiment.

For the experimentation, 30 random phrases were selected from the different reviews, and of these we chose those that when analyzed with both procedures allowed to obtain a positive or negative polarity, and discarded those phrases that in Sentistrength gave neutral results, since that would have prevented their comparison between the two procedures.

The following code in Python language was used to obtain the polarity of the texts analyzed in PMI:

```
from __future__ import division
import urllib.request
import urllib
import json
```

```

from math import log
def hits(word1,word2=""):
    query="http://ajax.googleapis.com/ajax/services/search/web?v=1.0&q=%s"
    if word2 == "":
        results = urllib.request.urlopen(query % word1)
    else:
        q = urllib.request.quote(word1+" "+"AROUND(10)"+ " "+word2)
        #print(q)
        #print (query % q)
        results = urllib.request.urlopen(query % q)
    json_res = json.loads(results.readall().decode('utf-8'))
    google_hits=int(json_res['responseData']['cursor']['estimatedResultCount'])
    return google_hits
def so(phrase):
    num = hits(phrase,"excellent")
    print (num)
    den = hits(phrase,"poor")
    print (den)
    ratio = num / den
    #print ratio
    sop = log(ratio)
    return sop
print (so("THIS is the place to stay at when visiting the historical area of Seattle"))

```

This code prepares a query to be made in the Google API, this query will be appended to the sentence we want to evaluate, and based on the result obtained after the query the code allows to display a value associated with ‘excellent’, ‘poor’ and ratio; this numerical value is important, since it allows to contrast it with the value obtained at the other end of the polarity, obtaining a result because it points to which end of the polarity this value belongs.

In addition to this procedure, we have submitted the same data set to Sentistrength. This classification method allows the evaluation of the existing polarity of a text. For this, the procedure assigns a value of 1 to 5 of positive and negative emotion load, being 1 a reduced emotional load and 5 a high emotional load. For example, if the value obtained in a sentence results in 2.5, we can infer that it has a slight positive charge and a strong negative charge. In that case the phrase would have a negative overall emotional load.

Once the results of the two binary classifiers have been obtained, a contingency table is elaborated with the objective of carrying out a ROC analysis, which allows us to identify the classifier that produces the results.

## EXPERIMENTATION

For this stage of the experimentation, we worked with a sample of phrases from Trip Advisor's online review site, in order to test the methodology applied.

**Table 1** shows the results of the analysis carried out with positive and negative data, which makes it possible to typify the emotion described.

Where:

**Real:** Values that were assigned according to the observation, analysis and interpretation of the text carried out by the experts and correlated to the polarity obtained as: values "+" Positive and "-" Negative.

**PMI:** Values obtained by means of binary classifier, "+" corresponds to Excellent and "-" corresponds to poor.

**SentiStrength:** Values obtained by binary classifier, and can take values "+", Positive and "-", Negative.

**Figure 1** shows the three points analyzed in the ROC space, where the point at the top left of the ROC curve is Polarity-Emotion, ranking results better than random; this implies that this classifier is the best possible method of prediction; while the points of the PMI classifiers, Sentistrength, are below the diagonal of the ROC curve, representing poor results (worse than random).

**Table 1.** Contingency Table

		Real		Sensibility	Ratio
		+	-	VPR	FPR
Classifiers	<b>Excellent (+)</b>	12	2	0,92	1
	<b>Poor (-)</b>	1	0		
	<b>Positivo</b>	12	1	0,92	0,5
	<b>Negativo</b>	1	1		

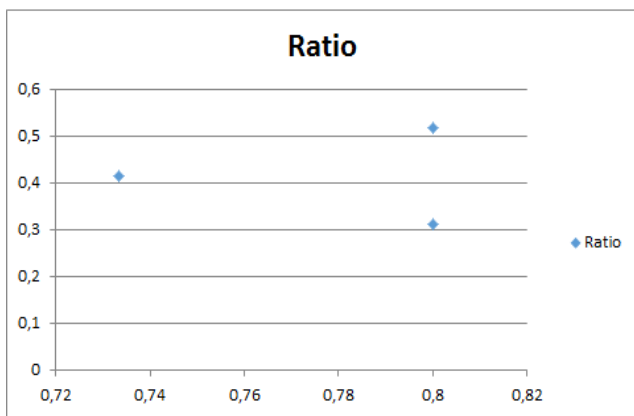


Figure 1. Ratio

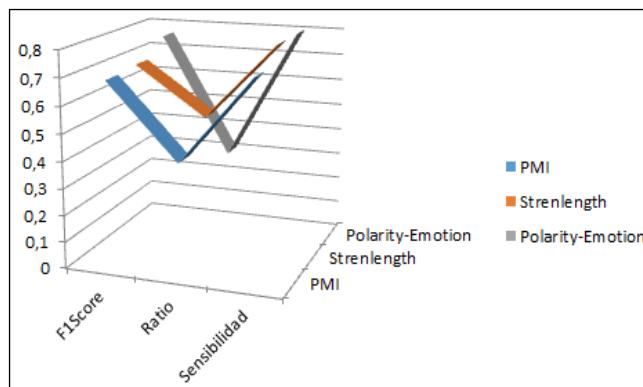


Figure 2. Ratio-Clasificadores

## CONCLUSIONS/RECOMMENDATIONS/SUMMARY

The results from the ROC analysis allow us to conclude that of the three classifiers used to evaluate the polarity detection effectiveness of hotel reviews, two are below the ROC curve, representing poor results considered to be worse than random. On the other hand, the third classifier is presented as a good predictor because it is located at the top left of the ROC curve, giving better than random classification results. In the future, other binary classifiers that demonstrate better performance in identifying the polarity in the text of hotel service reviews should be tested.

## REFERENCES

- Abirami, A., Askarunisa, A., Gayarathri, K. and Jeyalakshmi, M. (2015). Social Media Analysis for TamilNadu Tourism Places using VIKOR Approach. *Artificial Intelligent Systems and Machine Learning*, 7(5).
- Aurchana, P., Iyyappan, R. and Periyasamy, P. (2014). Sentiment Analysis in Tourism. *International Journal of Innovative Science, Enginnering & Technology*.
- Bakhtiyari, K. and Husain, H. (2014). Fuzzy Model on Human Emotions Recognition. *Recent Advances in Electrical and Computer Enginnering*, 77-82.
- Choo, E., Yu, T. and Chi, M. (2015). Detecting Opinion Spammer Groups through Community Discovery and Sentiment Analysis. *Lecture Notes in Computer Science*, 170-187. [https://doi.org/10.1007/978-3-319-20810-7\\_11](https://doi.org/10.1007/978-3-319-20810-7_11)
- Gomez, A. H. F., Bano, N. F. P., Martinez, C. C. E., Gómez, G. A. Á., Culque, T. W. V., Bustamante-Sanchez, N. S. and Sanchez, E. (2018). Introduction of the opinions of tourist services based on sentiment. *13th Iberian Conference on Information Systems and Technologies (CISTI)*, Caceres, 1-8. <https://doi.org/10.23919/CISTI.2018.8399376>
- Namahoot, C., Bruckner, M. and Panawong, N. (2015). Context-Aware Tourism Recommender System Using Temporal Ontology and Naive Bayes. *Advances in Intelligent Systems and Computing*, 183-194. [https://doi.org/10.1007/978-3-319-19024-2\\_19](https://doi.org/10.1007/978-3-319-19024-2_19)
- Prasath, R., Kumar, V. and Sudeshna, S. (2015). Assisting web document retrieval with topic identification in tourism domain. *Web Intelligence and Personalization on Social Media*, 31-41. <https://doi.org/10.3233/WEB-150308>
- Schickert, M., Liu, X. and Law, R. (2015). Hospitality and Tourism Online Reviews. *Travel & Tourism Marketing*.