I Know What You Felt Last Festival

Juandiego Morzán-Samamé
Universidad del Pacífico
Av. Salavrry 2020
Lima - Peru
Email: j.morzansamame@alum.up.edu.pe

Miguel Nunez-del-Prado
Universidad del Pacífico
Av. Salavrry 2020
Lima - Peru
Email: m.nunezdelpradoc@up.edu.pe

Hugo Alatrista-Salas
Universidad del Pacífico
Av. Salavrry 2020
Lima - Peru
Email: h.alatristas@up.edu.pe

***Abstract*—Festivals are an important leisure activity in the life of human beings. As a matter of fact, the organizers of festivals are interested in offering quality activities that allow them to position themselves in the entertainment market. To achieve this aim, the organizers use surveys to obtain a global opinion of the participants focusing on three key points: motivation, perception, and valuation. This method is tedious to perform and is time-consuming. In this effort, we present a complete process that enables to automatically obtaining an overall appreciation of a festival from tweets shared by participants. The aim of this contribution is to replace the surveys by the textual analysis of messages posted on social networks. The precision obtained in our experiments highlight the relevance of our proposal.**

Keywords—Sentiment analysis; machine learning algorithms; festival; survey analysis

I. Introduction

Festivals are recreative activities where people spend some time to have fun. People uses various sources of information to know in advance what kind of festivals may be interesting and offer activities according to their needs. In contrast, the organizers of festivals are interested in how to offer better conditions, events, etc. to grow up and to become more relevant over time. One of the most common methods to capture the information from the attendants is to apply surveys to participants. In this way, a group of experts carefully prepare a set of questions focused on gathering the opinion of the assistants. Those questions try to capture the perception of people in a given festival, motivation, and valuation. In one hand, perception concerns how people recognize or appreciate the overall organization of the festival. On the other hand, motivation captures the reasons why people attend the festival. Finally, the valuation estimates the quality of the festival. To apply a survey, the organizers deploy a vast amount of materials and human resources. Besides, a significant amount of time is needed to conduct surveys, store and analyze data from them.

Conversely, sentiment analysis or opinion mining is a way to evaluate written language to determine if the expression is favorable, unfavorable, or neutral, and to what degree. Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain entities or their characteristics. These entities could be products, services, organizations, events, among others. Several textual sources are used to perform sentiment analysis task (e.g., Twitter, Facebook, Instagram, etc.). The applications of sentiment analysis are broad and powerful. The ability to extract insights from social data is a practice that is being widely adopted by organizations across the world.

In this effort, we develop a technique to replace classical surveys by textual opinions expressed on micro-blogging social networks. In detail, we use survey questions to build a corpus for applying sentiment analysis to know the sentiment toward some aspects of a festival. To achieve our goal, the data available on social networks will be used to infer people's opinion. More specifically, posts from Twitter[1] will be analyzed to classify if a post about a festival represents a positive,neutral or negative impression from one of the three key aspects: motivation, perception or valuation.

The present work is organized as follows. Section II describes documents found in literature associated with sentiment analysis. The dataset at our disposal is described in Section III. Later, in Section IV, we detail the pre-processing task. Sections V and VI describe the methodology, and the results respectively. This article ends with the conclusions and the future work, shown in Section VII

II. Related Works

In this section, we present related works on sentiment analysis methods, highlighting its application in the study of people's perception about a festival via tweets. From a technical point of view, Sentiment Analysis approaches could be classified as 1) lexical-based method; and 2) machine learning or classifier-based methods. There are also methods combining both approaches. In this section, we detail the classifier-based method.Works based on machine learning method uses supervised learning algorithms to assign a polarity to a textual document. In recent years, some authors concentrated on the study the accuracy of these algorithms. One of the first works found in the literature was the one presented by Pang *et al.* [1] In this paper, the authors highlight the difference between the topic-based classification and the sentiment analysis. Additionally, they show the difficulty of the last technique. More technically, the authors compare different algorithms, such as Naïve Bayes, Support Vector Machines (SVM) and Maximum Entropy (ME) and conclude that the best accuracy is obtained by the SVM algorithm. Finally, the corpus used for experimentation was obtained from the Internet Movie Database (IMDb).

In the same spirit, Xia *et al.* [4] use two types of feature sets: the part-of-speech based feature sets and the word-relation based feature sets. These two elements are combined with three

[1]https://dev.twitter.com/

algorithms: Nave Bayes, ME and SVM algorithms. These three combinations are called fixed combination, weighted combination and meta-classifier combination. The authors conducted experiments using a document-level polarity corpus from Amazon[2], containing product reviews about four categories: books, DVDs, electronics, and kitchen. Authors obtained the best score (88.65%) using the weighted SVM algorithm.

Recently, Singh and Husain [2] compared three algorithms: Nave Bayes, Multilayer Perceptron, and Support Vector Machine. The authors used two textual datasets: the first contains comments of movies (IMDb), and the other one review of products, like games, toys, etc., from amazon.com (Blitzer). The results obtained by the authors demonstrate that for these two corpora, the SVM algorithm achieves the best accuracy (81.15%).

Some works compare different algorithms on corpus written in various languages. For instance, Boiy and Moens [3] discuss the difficulty of performing sentiment analysis on data written in English, French, and Dutch. To overcome this problem, the authors uses cascade and aggregate learners, *i.e.*, the classifiers were combined in a pipelined cascaded way. Furthermore, to obtain good results, the authors uses active learning to reduce the number of examples to be labeled. Finally, Boiy and Moens perform the SVM, Multinomial Nave Bayes (MNB) and ME on corpus written in English, Dutch and French. The results show a better accuracy for MNB algorithm on corpus written in English.

Concerning the Spanish language, G. Sidorov *et al.* [8] perform a sentiment analysis task on Spanish corpus. The authors present an analysis of various parameter settings for three classifiers: Support Vector Machines, Nave Bayes, and Decision Trees. Further, the authors used n-grams of normalized words as features and observed the results of various combinations of positive, negative, neutral, and informative sets of classes. Like the previous works, the SVM algorithm obtains better results.

Later, Tuarob *et al.* [15] address the limitations posed by the traditional bag-of-words based methods. They propose to use heterogeneous features, such as N-gram, Dictionary based compound features, Topic distribution features, Sentiment features and Combined features in combination with ensemble machine learning techniques like: Majority voting, Weighted Probability Average, Multi Staging and Reverse Multi Staging to discover health-related information. Several tests were run by the authors in order to determine which parameter configuration and machine learning algorithm suited the best for a feature set. In addition, they combined several classifiers trained with the previously obtained features set and compare them in terms of precision, recall and F1 measure.

Conversely, regarding the classic techniques to evaluate the quality of a festival, some papers are describing several methods [5], [7]. There exist others contributions in which, textual data describing touristic places, was used to contribute to know more about a tourism destination [6]. Nevertheless, to our knowledge, no documents uses the information captured by the surveys combining with analysis sentiment techniques.

[2]www.amazon.com

III. Datasets

In the present section, we describe the survey and micro-blog comments from Twitter datasets used for the present effort. The former dataset is composed of questions oriented to gather people opinion about motivation, perception or valuation. For a sake of clarity, we present an example of questions in Table I. The survey has 46 question, where 18 are about Motivation, 25 Perception and 3 Valuation.

TABLE I. Characteristics of Three Types of Vectors

Variable	Question
Motivation	I like to try foods and drinks of different cultures
Perception	The ticket price is right
valuation	I would recommend to other people attending this festival

The latter dataset contains gathered tweets from a Peruvian well known food festival named “mistura”. Thus, we used the Twitter API to collect tweets containing the hashtag *#mistura2016* or the word *mistura* within the bounding box of Peru during August and September 2016. Our tweet dataset has 15057 tweets. The captured features are described in Table II.

TABLE II. Characteristics of Four Types of Vectors

Variable	Description
Id user	User identifier
Id tweet	Tweet identifier
Datetime	Datetime of the emitted tweet
Message	Text and or emoticons containing an opinion
Language	The language of the tweet
Country	The country where the tweet took place

It is worth noting that we have drop the *Id tweet* and the *Id user* for privacy reasons. In the next section, we present how to use this information.

IV. Pre-Processing

Both the surveys and the tweets were pre-processed, differently, for later use. Accordingly, the process which followed for each one of them will be described below.

A. Pre-processing surveys

In this study, the relevance of the surveys is to form vectors corresponding to each class (motivation, perception and valuation) that contain the most important words associated with them and to help labeling the tweets. The procedure developed in order to pre-process the surveys went as follows:

1) **Sorting questions.** The questions were sorted according to the class (motivation, perception or valuation) to which they corresponded. In other words, the number of documents that compose the corpus of the surveys was reduced to three: the number of classes held.
2) **Converting the question string to lowercase.** This step enabled the further comparison with the stop-words dictionary and the application of the TF-IDF technique.
3) **Removing stopwords.** By comparing each word in the vectors with the provided dictionary by the NLTK

library[3] and the non-alphabetic characters. The result were relatively important space-separated words, for each question.

4) **Tokenization.** The questions' strings were tokenized and a vector was formed for each one of them containing their respective tokens.
5) **Lemmatization.** This was the final step. Using the Pattern.es tool from CLiPS[4], the tokens were finally tagged, leaving only the lemma of the original.

B. Pre-processing tweets

About the tweets, they will be used to be tagged and to check how accurately one can identify or classify each one of them in one of the three classes mentioned before: motivation, perception, and valuation. The process followed to pre-process the tweets is more extensive than the one developed in the case of the surveys. The reason behind is that the tweets text is more exposed and vulnerable to spelling mistakes, the use of slang and emoticons, the use of user mentions and hashtags, the use of non-alphanumeric characters, the style of the authors, among others. The procedure developed to pre-process the tweets went as follows:

1) **Tweets segmentation.** In this step, first, the language and country of origin of each tweet were identified. Then, the tweets written in Spanish language and on Peru were selected. The above served as to be more certain that the tweet was actually about the Peruvian gastronomic fair "Mistura" and avoid the problem of labeling and classifying tweets in another language, in which case probably the result would be a tweet without a class. The final corpus used for this task is equivalent to one-third of the total number of tweets collected.
2) **Tweet string conversion to lowercase.** This step is performed for the same purposes as in the case of the surveys, explained above.
3) **Elimination and standardization of special characters.** Through the encoding and decoding of the words under the UTF-8 framework, it was possible to remove special characters from the Spanish language such as cases of accent mark, umlaut, the Spanish letter "ñ", among others. This is achieved during the encoding and decoding process; those processes translate Unicode strings into the closest ASCII value and helps standardize the characters, which permits a better word to word comparison.
4) **Hashtag elimination.** user mentions, and URLs. The purpose of this study is not to follow nor identify tendencies, but to focus on the content of the tweet to achieve the labeling and classification objectives. That's the reason why it was decided to remove the tweet's elements that didn't add value to the content (like #s, user's mentions, and URLs).
5) **Stopwords and non-alphabetic characters elimination.** Each word on the tweet is compared to the NLTK's dictionary, and the stopwords were removed. The non-alphabetic characters, such as punctuation marks, were removed too. The result was a string of words separated only by a space.
6) **Repeated letters corrections.** In the case of finding a word containing a sequence of repeated letters greater than two, said sequence was reduced to a maximum of two. This helps mitigate drawbacks with the writing style of users (especially on social networks it is common to write words with excess letters).
7) **TF-IDF application.** Likewise, using the TF-IDF scoring, a specific dictionary of stopwords to the set of tweets. For this, those words that, in general, reported a lower TF-IDF score were eliminated. A clear example is "rt", which refers to "retweet" and is present in all tweets of this type.
8) **Tokenization.** Each tweet was tokenized, and a vector of tokens was created for each formed by the words.
9) **Lemmatization.** As with the surveys, using the Pattern.es tool of CLiPS, the tokens were tagged.

[3]NLTK library: http://www.nltk.org
[4]http://www.clips.ua.ac.be/pages/pattern-es

V. METHODOLOGY

The methodology followed in this study has a twofold objective. First, identify the most representative words to label documents talking about motivation, perception, and valuation. Then, use Machine Learning techniques to classify micro-blogging comments belong to one of the aforementioned classes. The overall process is detailed in the following sections.

A. Corpus Labeling

The corpus labeling has two steps. The first is the dictionary creation and the second one is the corpus labeling. To build the dictionary, we rely on the survey questions, which were elaborated by a specialist. Questions were formulated to gather people opinion about *motivation*, *perception* and *valuation*. Thus, we computed the *TF-IDF* over the tokenized surveys to obtain the ten more representative word for each class. Once the most representative words were obtained, we use them a weighted vector dictionary.

Since from the survey questions, we have three different vectors containing the most representative word used to talk about *motivation*, *perception* and *valuation*. We built some variants of these vectors to improve the classification of tweets. Consequently, we form four different dictionary vector as shown in Table III.

- **Type A vectors**. The original vectors, containing the ten most representative words of each class.
- **Type B vectors**. The original vector, additionally containing synonyms for each word. Within each of the three original vectors, one vector per word was formed, i.e., each new vector was composed of the word and its synonyms to keep the appropriate weights.
- **Type C vectors**. The original vectors, additionally containing conjugations of verbs. Within each of the three original vectors, one vector per word was created. In this case, each new vector was composed of

TABLE III. EXAMPLE OF CHARACTERISTICS OF THE MOTIVATION WEIGHTED DICTIONARY VECTORS

Type of Vector	Motivation dictionary vector content				
Weight	10	9	...	2	1
Type A	querer	disfrutar	...	experiencia	tener
Type B	querer	disfrutar	...	experimentar	tener
	desear	gozar	...	notar	haber
	...	...	...	...	...
	anhelar	gustar	...	percibir	poseer
Type C	querer	disfrutar	...	experimentar	tener
	quiero	disfruto	...	experimento	tengo
	...	...	...	...	...
	quisieron	disfrutan	...	experimentan	tienen
Type D	querer	disfrutar	...	experimentar	tener
	quiero	disfruto	...	experimento	tengo
	...	...	...	...	...
	quisieron	disfrutan	...	experimentan	tienen
	desear	gozar	...	notar	haber
	deseo	gozo	...	noto	tengo
	...	...	...	...	...

the word and its conjugation to maintain the appropriate weights.

- **Type D vectors**. The original vectors, additionally containing synonyms for each word, conjugations of the words that were verbs, and conjugations of the synonyms that were verbs. Within each of the three original vectors, a new vector was created per word. Each one contained the word, its synonyms, and conjugations.

The second step takes as input a set of weighted dictionaries vectors containing the terms, weight and TF-IDF value for motivation Dic_{mot}, perception Dic_{per} and valuation Dic_{val}, and a tweet *t* containing words $t = \{w_1, w_2, \ldots, w_n\}$. Hence, to establish whether t_i represents motivation, perception or valuation, the algorithm computes a scores based on the sum the weights given by a weighted vector dictionary.

$$
\begin{aligned}
motivation_{score}(t_i) = & \sum_0^j Dic_{mot}(w_{ij}) \\
perception_{score}(t_i) = & \sum_0^j Dic_{per}(w_{ij}) \\
valuation_{score}(t_i) = & \sum_0^j Dic_{val}(w_{ij})
\end{aligned}
$$

Finally, to label a t_i tweet the algorithm takes the highest score (*i.e.,* $max(motivation_{score}, perception_{score}, valuation_{score})$) to identify what the tweet is talking about (motivation, perception or valuation). In the case of a tie, and values different from zero, the algorithm is repeated using the TF-IDF score as weight as tiebreaker. If all the scores are equal to zero, then the t_i tweet is label as "No class".

TABLE IV. CHARACTERISTICS OF FOUR TYPES OF VECTORS

Type of Vector	A	B	C	D
#labeled tweets	5029	9189	5319	10020
#non-labeled tweets	10028	5868	9738	5037
Ratio of labeled tweets	0.33	0.61	0.35	0.67
Total amount of tweets	15057	15057	15057	15057

Using this technique, the algorithm was able to label 67% of the raw tweets, when using the *Type vector D* as shown in Table IV. This set of labeled tweets will become the corpus to label the reaming 33% of tweets. We detail this process in the following subsection.

B. Tweets Classification Results

Concerning the classification task, different Machine Learning techniques were used, such as *Logistic Regression* [9], *Naïve Bayes* [10], *Support Vector Machines* [11] (*SVM*), *Decision Trees* [12], *Random Forests* [13], and *Neural Networks* [14]. First, we construct the train set from the labeled tweets and the test set from the non-labeled tweets. Later, tweets are put in a vector and featured using the TF (term frequency) scoring and setting a maximum of 100 features. Finally, we apply the Machine Learning techniques to classify the train set. In order to evaluate the classifiers accuracy, a random sampling process was performed by setting and initializing an internal random number generator (random seed), which decided the (randomly) splitting of labeled tweets set into train and test subsets. Results, in terms of precision, recall and F-measure, are shown in Table V, where "M" stands for motivation, "P" for perception, and "V" for valuation.

As we can notice in Table V the better results were obtained using the SVM algorithm and for corpus belonging the Type A vector. This behavior is similar from those described in Section II. Once the tweets are classified, it is possible to perform other Text Mining techniques such as sentiment analysis, topic modeling, *etc...* to explore insights from people comments.

VI. DISCUSSION OF THE RESULTS

In this section, two results are described: the labeling and classification results. In one hand, concerning the labeling, the Table IV shows, for each type of vectors, the number of labeled tweets and their ratio *w.r.t.,* the total amount of tweets. On the other hand, regarding the classification task, the results of the different classifiers and the impact of the type of vector are shown in Table V.

A. Labeling

We can notice that the Type A vector is the one with the least amount of successfully labeled tweets. The reason behind this behavior it that it only considers the ten most representative words of each class. That is the main reason why other variations were incorporated to increase the number of labeled tweets (B, C, and D types). Other types maximum the number of tweets because they combine synonyms as well as the conjugations of the representative words. Finally, regarding the individual level, Type B vector increases the number of labeled tweets in greater measure than the Type C vector. In other words, the addition of synonyms into the original vectors generates a greater impact than the incorporation of only conjugations.

B. Classification

It is evident that the classification is much more accurate when using Type A and Type C vectors reaching even a 98% of accuracy. The reason is that both types of vectors contain a less amount of labeled tweets (almost the half of the other types). Regarding Type B and Type D vectors, there is no much difference between their classification accuracy. Nevertheless, there is a faint difference between them: Type D vectors works with synonyms, as well as with conjugations and, thus, achieves a greater amount of labeled tweets and a broader classification. Finally, it is important to notice that the

TABLE V. QUALITY MEASURES PER ALGORITHM AND PER VECTOR TYPE

		Type A vector			Type B vector			Type C vector			Type D vector		
Classifier	**Class**	**Accuracy**	**Recall**	**F-1 score**	**Accuracy**	**Recall**	**F-1 score**	**Accuracy**	**Recall**	**F-1 score**	**Accuracy**	**Recall**	**F-1 score**
Logistic Regression	M	0.99	1.00	0.99	0.89	0.91	0.90	0.98	0.93	0.95	0.88	0.94	0.91
	P	0.98	0.99	0.98	0.66	0.61	0.64	0.94	0.99	0.97	0.70	0.61	0.66
	V	0.95	0.78	0.86	0.67	0.59	0.63	0.97	0.83	0.90	0.88	0.32	0.47
	Avg/total	0.98	0.98	0.98	0.84	0.85	0.84	0.96	0.96	0.96	0.85	0.85	0.84
Naïve Bayes	M	0.98	0.97	0.98	0.88	0.93	0.90	0.95	0.93	0.94	0.85	0.96	0.90
	P	0.96	0.98	0.97	0.67	0.54	0.60	0.94	0.97	0.96	0.72	0.45	0.55
	V	0.88	0.76	0.82	0.68	0.57	0.62	0.93	0.83	0.88	0.92	0.33	0.49
	Avg/total	0.96	0.96	0.96	0.83	0.84	0.83	0.95	0.95	0.95	0.83	0.83	0.82
SVM	M	0.98	1.00	0.99	0.89	0.91	0.90	0.97	0.93	0.95	0.88	0.93	0.91
	P	0.98	0.98	0.98	0.65	0.62	0.64	0.95	0.98	0.96	0.70	0.64	0.67
	V	0.95	0.84	0.89	0.66	0.59	0.62	0.91	0.83	0.87	0.89	0.35	0.50
	Avg/total	0.98	0.98	0.98	0.84	0.84	0.84	0.95	0.95	0.95	0.85	0.85	0.85
Decision Trees	M	0.99	0.99	0.99	0.91	0.90	0.90	0.95	0.95	0.95	0.89	0.93	0.91
	P	0.97	0.98	0.98	0.64	0.66	0.65	0.96	0.96	0.96	0.71	0.64	0.67
	V	0.87	0.80	0.83	0.67	0.65	0.66	0.90	0.91	0.91	0.63	0.45	0.53
	Avg/total	0.97	0.97	0.97	0.85	0.85	0.85	0.95	0.95	0.95	0.85	0.85	0.85
Random Forests	M	0.96	0.99	0.98	0.91	0.92	0.91	0.94	0.94	0.94	0.89	0.94	0.91
	P	0.98	0.97	0.98	0.68	0.65	0.67	0.95	0.96	0.96	0.74	0.65	0.70
	V	0.91	0.80	0.85	0.72	0.87	0.69	0.87	0.83	0.85	0.67	0.41	0.50
	Avg/total	0.97	0.97	0.97	0.86	0.86	0.86	0.94	0.94	0.94	0.85	0.86	0.86
Neural Networks	M	0.99	1.00	0.99	0.91	0.92	0.91	0.98	0.94	0.96	0.89	0.94	0.92
	P	0.97	0.99	0.98	0.68	0.69	0.69	0.96	0.98	0.97	0.74	0.67	0.70
	V	0.88	0.76	0.84	0.71	0.59	0.64	0.93	0.91	0.92	0.76	0.42	0.54
	Avg/total	0.98	0.98	0.98	0.86	0.86	0.86	0.96	0.96	0.96	0.86	0.87	0.86

Neural Networks algorithm outperformed the other techniques and achieved an accuracy of $86-87\%$ when using the Type B and Type D vectors. Additionally, the Naïve Bayes classifier got the least accuracy, approximately 3% behind the Neural Networks.

VII. CONCLUSION

Surveys are a very powerful tool to collect the opinion about a recreational activity. Specifically, at festivals, surveys are designed to capture three main indicators: motivation, perception, and valuation. Even if surveys are an interesting tool, the process of data collection and the analysis is tedious and time-consuming. In this study, we present a process to obtain automatically an overall perception of the activity from posts shared by the participants in the social network, specifically, Twitter. In a first stage, a corpus of Tweets has built. Subsequently, we used the surveys to label our corpus. Finally, we use six algorithms of automatic learning to be able to predict the polarity of a tweet concerning one of the three indicators earlier mentioned. These results give us a global appreciation of the participants to a festival. To meet this goal, we had to face two main problems: 1) the difficulty of analyzing text in Spanish; and 2) the construction of a labeled corpus from a set of tweets and surveys. Another challenge that is no less important is the classification process on a corpus comprising unbalanced classes.

As future work, we wish to extend our proposal to messages in other languages, such as French and English, and from other social networks, such as Instagram or Facebook. Additionally, we want to include another kind of information such as meteorological data, and traffic data, to obtain richer results. Finally, some points can be improved in our proposals, such as the analysis of sarcasm or the treatment of ambiguity.

REFERENCES

[1] B. Pang, L. Lee, S. Vaithyanathan. *Thumbs up? Sentiment Classification using Machine Learning Techniques*. Proceedings of the conference on Empirical methods in natural language processing. 2002.

[2] P. Singh, M. Husain. *Methodological study of opinion mining and sentiment analysis techniques*. International Journal on Soft Computing, 5, 11, 2014.

[3] E. Boiy, M. F. Moens. *A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts*. Information Retrieval Journal, 12, 526. 2009.

[4] R. Xia, C. Zong, S. Li. *Ensemble of feature sets and classification algorithms for sentiment classification*. Information Sciences, 181, 1138. 2011.

[5] K. Small, D. Edwards, L. Sheridan. *A Flexible Framework For Evaluating The Socio-Cultural Impacts Of A (Small) Festival*. International Journal of Event Management Research. Volume 1, Number 1, 2005.

[6] A. Gassiot; L. Coromina. *Destination image of Girona: an online text-mining approach*. International Journal of Management Cases . 2013, Vol. 15 Issue 4, p301-314. 2013.

[7] S. Hudson, R. Hudson. *Engaging with consumers using social media: a case study of music festivals*. International Journal of Event and Festival Management, Vol. 4 Issue: 3, pp.206-223. 2013.

[8] G. Sidorov, S. Miranda-Jiménez, F. Viveros-Jiménez, A. Gelbukh, N. Castro-Sánchez, F. Velásquez, I. Díaz-Rangel, S. Suárez-Guerra, A. Treviño and J. Gordon. *Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets*. Advances in Artificial Intelligence. MICAI 2012.

[9] D. Hosmer, S. Lemeshow, R., Sturdivant. *Applied logistic regression* (Vol. 398). John Wiley and Sons. 2013.

[10] K. Murphy, *Naïve Bayes Classifiers*. University of British Columbia. 2006.

[11] M. Hearst, S. Dumais, E. Osuna, J. Platt, B. Scholkopf. *Support vector machines*. IEEE Intelligent Systems and their Applications, 13(4), 18-28. 1998.

[12] J. Quinlan, *Induction of decision trees*. Machine learning, 1(1), 81-106. 1986.

[13] L. Breiman, *Random forests. Machine learning*, 45(1), 5-32. 2001.

[14] D. Specht, *Probabilistic neural networks*. Neural networks, 3(1), 109-118. 1990.

[15] Suppawong Tuarob, Conrad S. Tucker, Marcel Salathe, Nilam Ram, An ensemble heterogeneous classification methodology for discovering health-related knowledge in social media messages, Journal of Biomedical Informatics, Volume 49, 2014, Pages 255-268.