Automatic Cyberbullying Detection in Spanish-language Social Networks using Sentiment Analysis Techniques

Rolfy Nixon Montufar Mercado, Hernan Faustino Chacca Chuctaya, Eveling Gloria Castro Gutierrez
National University of San Agustin
CiTeSoft
Arequipa-Peru

***Abstract*—Cyberbullying is a growing problem in our society that can bring fatal consequences and can be presented in digital text for example at online social networks. Nowadays there is a wide variety of works focused on the detection of digital texts in the English language, however in the Spanish language there are few studies that address this issue. This paper aims to detect this cybernetic harassment in social networks, in Spanish language. Sentiment analysis techniques are used, such as bag of words, elimination of signs and numbers, tokenization and stemming, as well as a Bayesian classifier. The data used for the training of the Bayesian classifier were obtained from the Spanish Dictionary of Affect in Language (SDAL), which is a database formed by more than 2500 words manually evaluated in three affective dimensions: Pleasantness, activation and imagery, as well as same 595 words obtained following the same procedure of SDAL was used with the help of the members of the Research Center, Technology Transfer and Software Development. As a result, the software developed has 93% success in the validation tests carried out.**

***Keywords*—Cyberbullying; social media analytics; sentiment analysis; tokenization; stemming; bag of words**

I. Introduction

As online social networks (OSN) have grown in popularity, instances of cyberbullying at OSN have become a growing concern. The prevalence of Cyberbullying in Peru is 20 to 40% in the last 10 years, according to the report "Cyberbullying: Approach to a comparative study: Latin America and Spain", by Albert Clemente, professor at the International University of Valencia (VIU) [1].

The VIU expert explains that in general, prevalence is understood as the set of individuals involved in the phenomenon of harassment or cyberbullying, that is, both victims, perpetrators and spectators. And stresses that "cyberbullying has not stopped growing and has become a problem in all cultures and regions of the world, both in its traditional and online" [1]. In addition, research has been conducted between technical performance tests and negative results, such as decreased school performance, absenteeism, school absenteeism, school dropout and violent behavior [2], and potentially psychological effects. devastating, such as depression, low self-esteem, suicidal ideation, and even suicide, which may have long-term effects on the future life of victims [3], [4]. Incidents of cyberbullying with extreme consequences, such as suicide, are reported routinely in the popular press.

Given the seriousness of the consequences that cyberbullying has on its victims and its rapid spread among college and university students, there is an immediate and compelling need for the research to understand how cyberbullying occurs in today's OSN. So things can be done to detect with cyberbullying.

The sentiment analysis, also called opinion mining [5], is the field of study that analyzes opinions, feelings, evaluations, attitudes and emotions of people towards entities such as products, services, organizations, individuals, problems, events, themes and their attributes. While most papers address it as a simple categorization problem, the sentiment analysis is actually a research problem [6] that requires addressing many natural language processing (NLP) tasks, including recognition of entities named [6], [7], the disambiguation of the polarity of the word [8], the personality recognition [9], the detection of sarcasm [10] and the extraction of the aspect [11]. In particular, the subtask is an extremely important subtask that, if ignored, the accuracy of the sentiment analysis in the presence of multiple points of opinion can be reduced consistently.

Therefore, the aspect-based sentiment analysis (ABSA) [6], [10], [12], [13], extends the feeling analysis section with a more realistic assumption that the polarity is associated with specific aspects (or characteristics of the product) instead of the whole text unit. For example, in the sentence "Food is delicious but service is horrible", the feeling expressed towards the two aspects is completely opposite. Through the aggregation of the analysis of feelings with the aspects, ABSA allows the model to produce a detailed opinion of the opinion of the people towards a particular product.

The objective sentiment classification (or objective-dependent) [14], [15], [16], instead, solves the polarity of the feeling of a given goal in its context, assuming that a prayer could express different opinions towards different specific entities. For example, in the sentence "I just logged into my Facebook and found an ugly picture of Anastacia", the sentiment expressed towards Anastacia is negative, while there is no clear feeling for Facebook.

Recently, Saeidi et al. [17], have tried to address the challenges of ABSA and the analysis of specific feelings. The task is to jointly detect the aspect category and resolve the polarity of the aspects with respect to a given objective. The deep learning methods [18], [19], [20], [21] have achieved great accuracy when applied to ABSA and analysis of specific

feelings. Especially, sequential neural models, such as short-term long memory networks (LSTM) [22], are of increasing interest for their ability to represent sequential information. In addition, most of these sequence-based methods incorporate the attention mechanism, which is rooted in the alignment model of machine translation [23]. Such a mechanism takes an external memory and representations of a sequence as input and produces a probability distribution that quantifies the concerns in each position of the sequence.

Currently the industry around the sentiment analysis, increased its popularity due to the proliferation of commercial applications, offering many challenging problems becoming a very active research area with a broad domains offering a strong motivation for research and offering many challenging problems, which had not been studied before, such as processing information from social networks Facebook, Twitter, Instagram, blogs, wikis and other mass media online [24], [25], [26], which speed up the way of sharing private and/or intimate information through its platforms facilitating users to get in close contact with others without taking into account the dangers that these involve [5], [24], [27].

This type of communication can be dangerous and have serious consequences, because the post messages can contain some types of abusive or offensive content through which threats such as cyberbullying may emerge [24], [27]. In general, adults may be able to establish a secure line of communication and are often more aware of curiosity to explore new fields without the capacity of the dangers existing in social networks. Conversely, children or teenagers [24], [27], often have a misperception of threats and must weigh the potential risks of this communication.

The remaining part of the paper is organized as follows. A related works in this paper is explained in Sections 2, 3 and 4. Materials and Methods are described in Section 5. The results with the experiment settings is introduced in Section 6. Conclusions are presented in Section 7 and some future works are provided in Section 8.

II. Related Work

Dan Olweus [28], one of the leading specialists in the world in bullying, developed the first criterion to identify the specific form of bullying, when it was discovered that the phenomenon is associated with a high rate of suicide attempts among adolescents and defined a harassment situation as one in which ”a student is assaulted or becomes a victim if he is exposed, repeatedly and for a time, to negative actions carried out by another student or several of them”. For this author in the harassment there is a clear intention to harm the other, either physically or morally, in such a way that the intimidation is constant and persists over time. It is very remarkable the imbalance of forces between the aggressor and the victim, especially because the latter has difficulty overcoming mockery or aggression and decides to remain silent.

Vilares David [29], describes a system of opinion mining that classifies the polarity of texts in Spanish. He proposed an approach based on natural language processing that led to a segmentation, tokenization and labeling of the texts to then obtain the syntactic structure of the sentences by algorithms of dependency analysis.

The syntactic structure is then used to deal with three of the most significant linguistic constructions in the field we are dealing with: intensification, adversative subordinate clauses and denial. The experimental results show an improvement of the performance with respect to the purely lexical systems and reinforce the idea that the syntactic analysis is necessary to achieve a robust and reliable sentiment analysis.

Hernandez Li [30], carried out an investigation on the sentiment analysis in texts based on semantic approaches with linguistic rules for the classification of polarity of texts in Spanish, the classification was made according to a dictionary of semantic orientation where each The term is marked with a use value and emotional value, along with linguistic rules to solve several constructions that could affect the polarity of the text. For this evaluation a sample of 60,798 Twitter messages was used, each tweet is labeled with a global polarity, indicating whether the text expresses a Strongly Positive, Positive, Neutral, Negative, Strongly Negative feeling and no feeling. Among the results, it was found that 35.22% do not express any feelings, the 34.12% company positive feelings and 18.56% express negative feelings.

Martnez et. al and Alonso [31], [32], carried out a research approach to the study of the analysis of opinions in Spanish, where a survey of the researchs on the analysis of feelings is made and the small number of researches is expressed in Spanish, being the majority in English; also highlights the research in Spanish conducted by the group ITALICA of the University of Seville. Similarly, it tells us about the unbridled growth of the use of social networks where users give opinions of any type and topic, encouraging the use of these data for future research.

Baquero Abel [33], designed an instrument to detect cyberbullying in a school context and analyzed its psychometric properties. As participants, it had 299 adolescents (54.2% women and 45.8% men) with an average age of 15 years, belonging to the low stratum (22.1%) and middle stratum (78%). A quantitative study was carried out with a non experimental design of instrumental type and cross section. Under the classical theory of the tests, an adequate internal consistency was obtained, as well as convergent validity with the other measures.

The exploratory factor analysis was carried out in SPSS version 21, which yielded three factors. From the item response theory, it was found that the INFIT of the items ranged between 0.73 and 1.23 and the OUTFIT between 0.74 and 1.24. Based on the favorable results of the psychometric analysis, it is concluded that the instrument can be used for the detection of cyberbullying in a school context.

As instruments, the bullying prevention and dismantling project was used, which included bullying and cyberbullying questionnaires and workshops conducted by the school guidance team. Among the results revealed for the research, 58.32% have more of 200 Facebook contacts, a 21% share their password with pairs, and five students of the course answered having been bothered by this page.

Becerra Martn [34], analyzed the large volumes of data generated in social networks about public opinion and proposed to analyze a set of data using a sentiment classifier to tag publications made by Twitter users, in conjunction with

clustering algorithms for to be able to detect which are the topics on which opinions are expressed. He used a base of 2000 reviews of films labeled as positive and negative to then train an SVM classifier of feelings, then the K-Means clustering algorithm to get a general overview of the topics and an approximation of the feeling associated with them.

III. Sentiment Analysis

The sentiment analysis [30], [34], [35], seeks to extract opinions, about a certain entity and its different aspects from the natural language of texts. This is done automatically using algorithms for classification. Opinions are classified according to the feeling they transmit, that is, as positive, negative or neutral. Its importance is that our perception of reality, and thus also the decisions we make, is conditioned in a certain way by how other people see and perceive the world. That is why, from a point of view of utility, we want to know the opinions of other people on topics of interest, since they have various applications such as recommending products and services, determining which political candidate to vote in the next elections or even measuring public opinion before the measure taken by a company or a government.

A. Types of Sentiment Analysis

At the time of extracting this information, there is a great variety of methods and algorithms depending on the level of granularity of the analysis that we want to carry out. The levels [34], [36], [37], document, sentence or aspect are distinguished. The analysis at the document level determines the general feeling expressed in a text, while the analysis at the sentence level specifies it for each of the sentences in the text. However, these two types of analysis do not delve into in detail the element that people like or dislike. They do not specify what is the opinion, since considering the general opinion of an object as positive or negative does not mean that the author has a positive or negative opinion of all aspects of that object. For this work we focus on conducting a document level analysis as a first instance, due to the limit in the messages, the authors are usually concise and go straight to the point without having the possibility of including several different aspects in a single post. For this reason, using the post as a unit of analysis seems to provide an adequate level of granularity to make a broken down analysis of sentiment.

1) The sentiment analysis at the document level: The document-level analysis [34], [36], aims to classify the opinion of a document, in this case a post. This task does not consider the details regarding entities or aspects, but considers the document as a whole, which will be labeled as positive or negative. This can be considered as a traditional text classification task, where classes are different orientations in terms of feelings. However, to ensure that this type of analysis makes sense, we assume that each document expresses a single opinion on a single entity. Although this may seem a limitation, because in a post one could express more than one opinion towards different entities, in practice it produces positive results, since users tend to focus on only one aspect in each post.

IV. Cyberbullying

Cyberbullying, [24], [27], [38], [39], is the use of digital media to harass a person or group of people, through personal attacks, disclosure of confidential or false information among other means. It may constitute a criminal offense. Cyberbullying involves recurrent and repetitive damage inflicted through digital media.

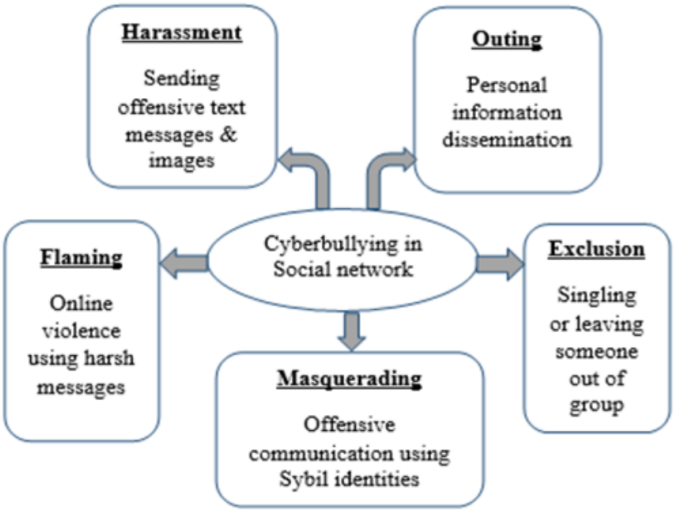

Fig. 1. Types of cyberbullying (Source: Hosseinmardi [27]).

According to Karthik Dinakar, [40], [3], cyberbullying is a more persistent version of traditional forms of intimidation, which extend beyond the physical confines of a school, sports field or workplace, with the victim often does not experience any respite from it. Cyberbullying gives an individual the power to embarrass or hurt a victim before an entire online community [41], especially in the realm of social networking websites. This is widely recognized as a serious social problem, [38], [40], [3], [42], [43], especially for teenagers.

The mitigation of cyberbullying involves two key components, robust techniques for effective detection and reflective user interfaces that encourage users to reflect on their behavior and choices. The types of cyberbullying usually occurs in the social network that shows in Fig. 1 are Harassment (sending offensive text messages and images), Flaming (Online violence using harsh messages), Masquerading (Someone might create fake email addresses or instant messaging names or someone might use someone else's email or mobile phone to bully another person), Outing (personal information dissemination) and Exclusion (Singling or leaving someone out of group) [27].

V. Materials and Methods

A. Database

As a first step for the detection of cyberbullying through the analysis of feelings, it is necessary to have a database for the training of the Bayesian network. The database of Agustn Gravano (SDAL) [25], from the Faculty of Exact and Natural Sciences of the University in Buenos Aires, Argentina, was used.

The database SDAL [25] is a lexicon of 2880 words in Spanish, which have been annotated manually with respect to three affective dimensions:

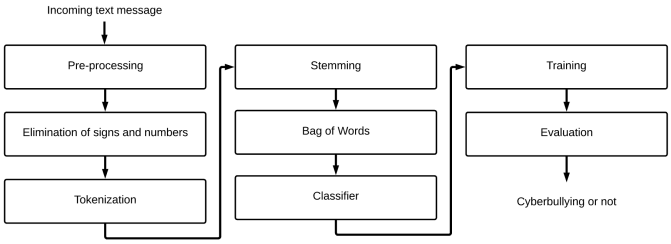

Fig. 2. Cyberbullying detection process.

- Pleasant (pleasant, neutral, unpleasant)
- Activation (active, neutral, passive)
- Imaginability (easy to imagine, neutral, hard to imagine)

Likewise, 595 words obtained following the same SDAL procedure were used, collected with the help of the members of the Research Center, Technology Transfer and Development of Software (CiTeSoft), mostly Peruvian and Spanish slang in order to improve the results.

B. Cyberbullying Detection

For cyberbullying detection, the developed software follows the procedure shown in Fig. 2. Each process will be explained in greater detail in the following subsections.

C. Preprocessing

The message or post must be preprocessed because it contains unstructured text. The purpose of preprocessing is to transform messages into a uniform format that can be understood by the learning algorithm. In preprocessing, the process of tokenization, stemming, elimination and stoppage of meaningless words, elimination of numbers and blank spaces is carried out.

D. Bag of words

One of the most important subtasks in the text classification with bullying is the extraction of characteristics. Through the use of machine learning algorithms to train the classifier, the representation of the text as a feature vector is required. For that, a model commonly used in the processing of natural language is the Bag of Words (BoW) model. The main stage of this model is the creation of a vocabulary of words that, in our approach, indicates the vocabulary or the collection of abusive words. Among the reference approaches for text classification, the BoW approach has the highest recovery rate of 66% [44]. In the BoW model, each word is associated with a count of occurrences. This vocabulary can be understood as a set of non-redundant words where order does not matter. The BoW approach ignores grammar and detects offensive sentences by checking whether or not they contain offensive or offensive words.

E. Natural Language Processing

The stage of Processing of Natural Language [45], is very important for the implementation of models of analysis of feelings. It is necessary to carry out some processes both to the text that we are going to analyze, and to the text that the classifying algorithm will train. The processes that they applied are the following.

1) Elimination of signs and numbers: It is necessary to eliminate signs and numbers from the text, signs like “!”, “?”, “+”, etc., since the existence within the text could affect the recognition of the expressions by the classifier. Table I shows two examples.

TABLE I. EXAMPLE OF ELIMINATION OF SIGNS AND NUMBERS

Original text	Transformed Text
Que buena pelcula!	Que buena pelcula
Eres una mala persona :8	Eres una mala persona

2) Tokenization: It consists in breaking up the text in the different words of the ones that appear, naming these resulting elements tokens [46]. Each document in our corpus is transformed into a list of terms called tokens. This representation of data is also known as a bag of words. Tokens are strings of characters between spaces or punctuation, but this is not always the case, as for example in the case of the abbreviations [34]. The total set of words used, distinct and unique, is the vocabulary of the corpus. Table II shows two tokenization examples.

TABLE II. TOKENIZATION EXAMPLE

Original text	Transformed Text
Que buena pelcula	[”Que” ”buena” ”pelcula”]
Eres una mala persona	[”Eres” ”una” ”mala” ”persona”]

3) Stemming: It consists of extracting stems of the tokens obtained in the previous process. So the different forms, such as diminutives, superlatives, gender, etc. do not affect the result [47].

Stemming is the process of reducing inflected (or sometimes derived) words to their word stem, base or root form-generally a written word form. The stem need not be identical to the morphological root of the word; it is usually sufficient that related words map to the same stem, even if this stem is not in itself a valid root.

4) Naive Bayes classifier: Naive Bayes classifier, [26], assign probabilities to the data entered, building a tree of probabilities according to the data entered, within the NLTK tool set we find the nltk. Naive Bayes Classifier class that allows us to use this type of classifier and train it according to our needs.

5) Training: To train the Naive Bayes classifier we need known data, so it is a supervised learning algorithm. This is where the need for a lexicon arises because the analysis was based on them. The lexicon is a file that can vary in its structure, but it must contain a list of words, with its respective subjectivity value in order to be processed in order to train the classifying algorithm. In this case we are using the database SDAL [25], which we saw in detail in the database section.

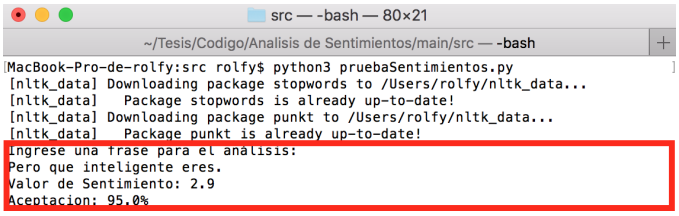

Fig. 3. Example of text with positive polarity (low probability bullying).

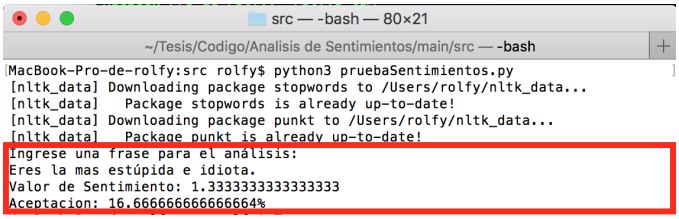

Fig. 4. Example of text with negative polarity (high probability of nullying).

6) Implementation: It was imported and used:

- NLTK [48] as a Python library for natural language processing.
- Pickle to save the classifier instance as a binary file.
- OS to be able to interact with the system.

The classifiers require to receive a dictionary that recognizes them as features that are those that will describe conditions for a result to be given. In the particular case of our analysis we simply send the expression or word as a characteristic "word" the dictionary.

Then declare the global variables that correspond to the classifier and the list of words known by the classifier.

VI. Results and Validation

The software provides: the feeling value of the text or phrase which is in the range of 1 to 3, where 1 is negative, 3 is positive and 2 is neutral. In addition to acceptance, which is in the range of 0% to 100%, where 0% is a text or phrase has a high probability of containing bullying and 100% a low probability of containing bullying.

To validate the operation of the software a list of phrases was made, classified by three types: phrases and texts with positive polarity (high probability of bullying), negative (low probability of bullying) and neutral, which were evaluated by the software and confronted with the manual evaluation by the members of the CiTeSoft [49] of The National University of San Agustn [50]. Below are three types of phrases.

A. Without Bullying

The software successfully responded to the tests that were carried out with phrases and texts of positive polarity, as seen in Fig. 3 the phrase "But how intelligent you are" obtains an acceptance of 95 % indicating that there is a low probability of bullying in addition to indicating that it is a very positive phrase.

B. With Bullying

Tests were carried out with simple and complex negative polarity text, as we can see in Fig. 4, which is a container text of bullying. The software successfully responded to bullying text tests, as shown in the figure "You are the stupidest and idiot person." obtains an acceptance of 16% indicating a high probability of the existence of bullying and validating the operation of the software for the detection of text containing bullying.

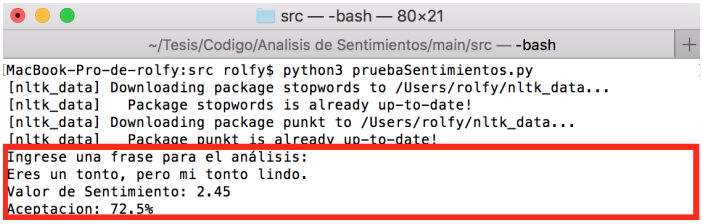

Fig. 5. Example of ambiguous phrase (neutral).

C. Ambiguous (Neutral)

Tests were performed with neutral polarity text, neutral polarity occurs when a text is ambiguous because in the first part the sentence contains a high probability of containing bullying but in the second part a low probability or vice versa.

The software successfully responded to tests with neutral text, as seen in Fig. 5 the phrase "You're a fool but my cute fool." The first part of the sentence has an insult, but in the second it is clarified that it is an expression of affection; obtaining an acceptance of 62% indicating a low probability of the existence of bullying and validating the operation of the software for the detection of ambiguous text.

D. Validation

To evaluate the effectiveness of the software, a collection of phrases from social networks (Facebook, Twitter, Instagram and Youtube) of diverse topics was done and a manual score was made between 0 and 10, where 0 represents a hurtful, offensive or bullying phrase and 10 a pleasant phrase or without bullying. This evaluation was carried out by the members of the CiTeSoft [49] (Center for Research, Technology Transfer and Software Development), then an arithmetic average was made between the evaluations of these members to be able to compare with the evaluation of the software developed. On the other hand, the evaluation of the same sentences by the software was made, then the range of acceptation that the software gives us from 0-100 to 1-10 was made to make a confrontation and see the effectiveness of it. As we observe in Table III, it shows the results of 100 sentences evaluated by members of CiTeSoft and the resulting average of each sentence. Likewise, the comparison between the average of the evaluations and the evaluation of the developed software, in version 1 and version 2, is shown in Table IV.

Finally, a comparative graph was drawn up as shown in Fig. 6 to see better the difference between the results obtained and the error percentage of the software. As can be seen in test 27, there was a very high error margin. This occurs because the software does not know the words that were used in the evaluation phrase.

TABLE III. RESULTS OF THE EVALUATION OF THE SENTENCES BY MEMBERS OF CITESOFT [49]

PHRASE	T 1	T 2		T 8	T 9	T 10	AVERAGE
1	1	5		6	3	5	3.8
2	1	3		1	1	1	1.8
3	7	5		7	6	5	6
4	9	7		9	8	9	7.8
5	1	5		5	1	5	3.1
6	9	6		9	8	8	8.1
7	1	5		3	2	3	2.4
8	1	2		1	2	1	2
9	1	4		2	2	1	1.8
10	1	3		3	2	2	2.3
11	9	6		9	7	9	7.6
12	1	2		3.5	3	4	2.5
13	9	7		8	8	9	7.9
14	1	2		2	2	2	1.7
15	9	6		9	9	9	8.4
16	1	2		1	1	1	1.6
17	1	2		1	1	1	1.7
18	1	3		6	3	6	3.4
19	1	3		3	2	2	2.4
20	8	6		9	9	9	8.5
21	1	3		1	2	1	1.6
22	9	7		8	8	8	7.4
23	9	6		9	7	9	7.7
24	1	4		3	4	3	2.6
25	1	3		1	1	1	1.6
26	1	5		3	1	3	2.3
27	1	5		1	2	1	2.6
100	1	4		3	4	3	3

TABLE IV. COMPARISON BETWEEN THE EVALUATION OF THE SOFTWARE (VERSION 1 AND 2) AND THE EVALUATION OF THE CITESOFT [49] MEMBERS OF THE TEST SENTENCES

PHRASE	AVERAGE	SOFTWARE V1	SOFTWARE V2
1	3.8	3.9	3.9
2	1.8	5	1.6
3	6	6.5	6.5
4	7.8	7.3	7.3
5	3.1	4	2.6
6	8.1	7.5	7.5
7	2.4	6.5	3
8	2	7.6	2.6
9	1.8	2.6	2.6
10	2.3	3.2	3.2
11	7.6	7.2	6.6
12	2.5	4	4
13	7.9	9.5	9.5
14	1.7	7.5	3.3
15	8.4	7.4	6.8
16	1.6	4.6	3.3
17	1.7	3.5	3.5
18	3.4	5.2	5.2
19	2.4	6	4.3
20	8.5	7.3	6.6
21	1.6	3.7	3.7
22	7.4	6.4	5.3
23	7.7	6.2	5.5
24	2.6	6.6	5
25	1.6	6.6	4
26	2.3	7.7	6.1
27	2.6	6.5	6.5
...	...	..	..
100	3	7.6	7.6

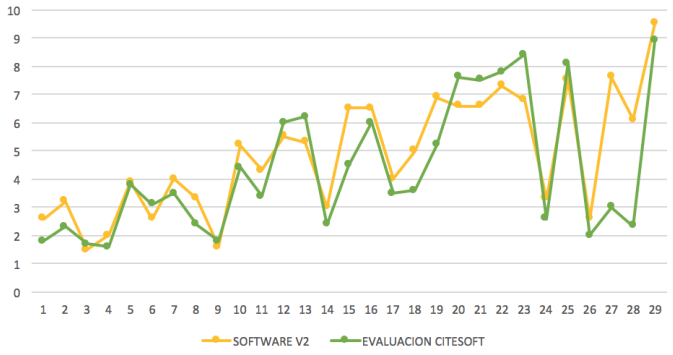

Fig. 6. Comparison between the evaluation of the test phrases of the software and the manual evaluation by the CiTeSoft members.

VII. CONCLUSION

In the validation of the software, three types of tests were carried out, without bullying, with bullying and ambiguous (neutral), then they were confronted with the manual evaluation of the members of the CiTeSoft [49], as shown in Table III, passing successfully the same in 93% of the cases, demonstrating its correct functioning.

The performance of the software developed depends directly on the number of words used in the sentences to be evaluated and if they are found in the word bag. In this work we worked with Peruvian and Spanish words and slang if we want to use software to evaluate phrases from other countries, we recommend adding words from these countries to the word exchange for better performance.

VIII. FUTURE WORK

As future work it is proposed to optimize the detection of cyberbullying by: Replacement of emoticons: A bag of emoticons and their respective meaning will be created, then this chain of characters will be replaced by a string that can be searched in the semantic orientation dictionary.

Correction of abbreviations: some of the most common abbreviated words will be replaced by their recognized grammatical form (Example:"q" $\rightarrow$ "que", "xq" $\rightarrow$ "$porque$").

Spelling correction: The Levenshtein algorithm with its notion of distance will be used. To correct the words, a dictionary of words will be used, which is made up of the complete list of forms of the Corpus of Reference of Actual Spanish (CREA) of the Royal Spanish Academy, with frequencies of use and with the conjugated forms most used , approximately 128 thousand forms. If a word is not found in the dictionary, the algorithm will take the nearest word with distance 1, and replace it with.

Correction of repeated characters: especially in the case of vowels, the repetition of the same concurrence will be replaced by a single one, with the exception of cc, rr, ll. Once the clean text is obtained, we proceed to carry out the lemmatization of the words to obtain their motto without conjugation, together with the tokenization and the segmentation of the sentences in order to classify the polarity (for example: "$largooooooo$" $\rightarrow$ "$largo$").

ACKNOWLEDGMENT

We would like to thank Citesoft UNSA Research Center for providing us materials, books, tools and computers to test algorithms proposed in this paper.

REFERENCES

[1] I. Viu, "Ciberacoso. Aproximación a un estudio comparado: Latinoamérica y España 1," pp. 1–28, 2015.

[2] R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R. Lattanner, "Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth," *Psychological Bulletin*, vol. 140, no. 4, pp. 1073–1137, 2014.

[3] K. Dinakar, R. Picard, and H. Lieberman, "Common sense reasoning for detection, prevention, and mitigation of cyberbullying," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2015-Janua, no. 3, pp. 4168–4172, 2015.

[4] M. van Geel, P. Vedder, and J. Tanilon, "Relationship Between Peer Victimization, Cyberbullying, and Suicide in Children and Adolescents," *JAMA Pediatrics*, vol. 168, p. 435, 5 2014.

[5] B. Liu, *Sentiment Analysis and Opinion Mining*. 2012.

[6] E. Cambria, D. Das, S. Bandyopadhyay, and A. F. Editors, "Socio-Affective Computing 5 A Practical Guide to Sentiment Analysis," 2017.

[7] Y. Ma, H. Peng, and E. Cambria, "Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM," 2014.

[8] Y. Xia, E. Cambria, b. Amir, H. @bullet, and H. Zhao, "Word Polarity Disambiguation Using Bayesian Model and Opinion-Level Features," 2015.

[9] Y. Ma, E. Cambria, and S. Gao, "Label Embedding for Zero-shot Fine-grained Named Entity Typing," pp. 171–180, 2016.

[10] E. Cambria, S. Poria, A. Gelbukh, I. P. Nacional, and M. Thelwall, "AFFECTIVE COMPUTING AND SENTIMENT ANALYSIS Sentiment Analysis Is a Big Suitcase," 2017.

[11] A. Mukherjee and B. Liu, "Aspect Extraction through Semi-Supervised Modeling," *Jeju, Republic of Korea*, pp. 339–348, 2012.

[12] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. María Jiménez-Zafra, and G. Eryiğit, "SemEval-2016 Task 5: Aspect Based Sentiment Analysis," pp. 19–30, 2016.

[13] E. Cambria, J. Fu, F. Bisio, and S. Poria, "AffectiveSpace 2: Enabling Affective Intuition for Concept-Level Sentiment Analysis," 2015.

[14] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective LSTMs for Target-Dependent Sentiment Classification," 2015.

[15] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification," pp. 49–54, 2014.

[16] B. Wang, M. Liakata, A. Zubiaga, and R. Procter, "TDParse: Multi-target-specific sentiment recognition on Twitter," vol. 1, pp. 483–493, 2017.

[17] M. Saeidi, G. Bouchard, M. Liakata, and S. Riedel, "SentiHood: Targeted Aspect Based Sentiment Analysis Dataset for Urban Neighbourhoods," 2016.

[18] T. H. Nguyen and K. Shirai, "PhraseRNN: Phrase Recursive Neural Network for Aspect-based Sentiment Analysis," pp. 2509–2514, 2015.

[19] Y. Wang, D. Zeng, B. Zhu, X. Zheng, and F. Wang, "Patterns of news dissemination through online news media: A case study in China," *Information Systems Frontiers*, vol. 16, no. 4, pp. 557–570, 2014.

[20] D. Tang, B. Qin, and T. Liu, "Aspect Level Sentiment Classification with Deep Memory Network," 2016.

[21] Y. Wang, M. Huang, L. Zhao, and X. Zhu, "Attention-based LSTM for Aspect-level Sentiment Classification," pp. 606–615, 2016.

[22] Hochreiter Sepp and Schmidhuber Jurgen, "Long Short-Term Memory," 2001.

[23] D. Bahdanau, K. Cho, and Y. Bengio, "NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE," 2015.

[24] K. B. Kansara and N. M. Shekokar, "A Framework for Cyberbullying Detection in Social Network," *International Journal of Current Engineering and Technology*, vol. 5, no. 1, pp. 494–498, 2015.

[25] D. A. Ríos and G. Matías, "A Spanish Dictionary of Affect in Language Spanish DAL : A Spanish Dictionary of Affect in Language," 2015.

[26] J. H. Xue and D. M. Titterington, "Comment on "on discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes"," *Neural Processing Letters*, vol. 28, no. 3, pp. 169–187, 2008.

[27] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra, "Detection of Cyberbullying Incidents on the Instagram Social Network," 2014.

[28] D. Olweus, "Bullying at School," pp. 97–130, 1994.

[29] D. Vilares, M. A. Alonso, and C. Gómez-Rodríguez, "Clasificación de polaridad en textos con opiniones en español mediante análisis sintáctico de dependencias Polarity classification of opinionated Spanish texts using dependency parsing," 2013.

[30] R. Hernández Petlachi and X. Li, "Análisis de sentimiento sobre textos en Español basado en aproximaciones semánticas con reglas linguísticas," *Tass 2014*, 2014.

[31] L. F. Hurtado and F. Pla, "ELiRF-UPV en TASS 2016: Análisis de sentimientos en twitter," *CEUR Workshop Proceedings*, vol. 1702, no. September, pp. 47–51, 2016.

[32] L.-F. Hurtado and F. Pla, "ELiRF-UPV en TASS 2014: Análisis de Sentimientos, Detección de Tópicos y Análisis de Sentimientos de Aspectos en Twitter," *Procesamiento del Lenguaje Natural*, pp. 1–7, 2014.

[33] C. A. Baquero and P. B. L. Avendaño, "Diseño y análisis psicométrico de un instrumento para detectar presencia de ciberbullying en un contexto escolar," *Psychology, Society, & Education*, vol. 7, no. 2, pp. 213–226, 2015.

[34] C. D. T. Estudiantiles, "Análisis de sentimientos en Twitter : El bueno , el malo y el ¿ :(," pp. 184–209, 2017.

[35] T. Wilson, J. Wiebe, and P. Hoffman, "Recognizing contextual polarity in phrase level sentiment analysis," *Acl*, vol. 7, no. 5, pp. 12–21, 2005.

[36] M. Hu and B. Liu, "Mining and summarizing customer reviews," *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04*, p. 168, 2004.

[37] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," *Proceedings of the 20th international conference on Computational Linguistics*, p. 1367, 2004.

[38] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. De Jong, "Improving cyberbullying detection with user context," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7814 LNCS, pp. 693–696, 2013.

[39] T. Nasukawa, T. Nasukawa, J. Yi, and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," *Proceedings of the 2nd international conference on Knowledge capture*, p. 70–77, 2003.

[40] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the Detection of Textual Cyberbullying," *Association for the Advancement of Artificial Intelligence*, pp. 11–17, 2011.

[41] N. Tsirakis, V. Poulopoulos, P. Tsantilas, and I. Varlamis, "Large scale opinion mining for social, news and blog data," *Journal of Systems and Software*, vol. 0, pp. –, 2016.

[42] M. Dadvar and F. D. Jong, "Cyberbullying detection: a step toward a safer Internet yard," *Proceedings of the 21st international conference ...*, pp. 121–125, 2012.

[43] M. Dadvar, F. M. G. de Jong, R. J. F. Ordelman, and R. B. Trieschnigg, "Improved cyberbullying detection using gender information," *12th Dutch-Belgian information retrieval workshop (DIR 2012)*, pp. 23–25, 2012.

[44] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, "Detecting offensive language in social media to protect adolescent online safety," *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*, pp. 71–80, 2012.

[45] K. Dinakar, R. Reichart, H. L. T. S. M. Web, and U. 2011, "Modeling the detection of Textual Cyberbullying.," *aaai.org*, 2011.

[46] C. Za'in, M. Pratama, E. Lughofer, and S. G. Anavatti, "Evolving Type 2 Web News Mining," *Applied Soft Computing*, pp. –, 2017.

[47] J. B. Lovins, "Development of a stemming algorithm," *Mechanical Translation and Computational Linguistics*, vol. 11, no. June, pp. 22–31, 1968.

[48] E. Loper and S. Bird, "NLTK: The Natural Language Toolkit," no. July, pp. 69–72, 2002.

[49] CiTeSoft, "CiTeSoft – Centro De Investigación, Transferencia De Tecnologías Y Desarrollo De Software I+D+I - UNSA."

[50] UNSA, "Universidad Nacional de San Agustin de Arequipa."