© International Journal of Combinatorial Optimization Problems and Informatics, Vol. 7, No. 2, May-Aug 2016, pp. 22-34. ISSN: 2007-1558.

Proposal of a Sentiment Analysis Model in Tweets for Improvement of the Teaching - Learning Process in the Classroom Using a Corpus of Subjectivity

Guadalupe Gutiérrez Esparza[1,2], Alejandro Padilla Díaz[3], Juana Canul-Reich*[1], Carlos Alejandro De-Luna[2], Julio Ponce[3].

Universidad Juárez Autónoma de Tabasco[1], Universidad Politécnica de Aguascalientes[2], Universidad Autónoma de Aguascalientes[3].

guadalupe.gutierrez@upa.edu.mx, apadilla2004@hotmail.com, juana.canul@ujat.mx, alejandro.deluna@upa.edu.mx, julk_cpg@hotmail.com

* *Corresponding Author*

Abstract. In this paper, we propose a sentiment analysis model for the assessment of teacher performance in the classroom by tweets written by a pilot group of college students. Naive Bayes (NB) is the technique to be applied to classify tweets based on the polar express emotion (positive, negative and neutral), to carry out this process, a dataset fits adding distinctive terms of context as possible features to support the classification process.

Keywords: Sentiment Analysis, Naive Bayes, Corpus of Subjectivity, Twitter

1 Introduction

Since long time ago, the human being had used the media to express their needs, preferences and emotions. The Internet has been one of the most used media to make the communication possible and find information of interest. The growth of social networks like: Twitter, Facebook, LinkedIn, Google+, Pinterest and others, has generated a large amount of information about the preferences and behavior of the users. The majority of the data that is constantly generated in the social networks could contain valuable information like perceptions and tendencies from the users to the objects, personalities or services. To process the data that exist in the social networks is necessary the usage of linguistic mechanisms that contribute to define the correct sense of the words. One of the areas that have surged of the Natural Language Processing (NLP) to analyses the information contained in the web either in sentences or in documents is the sentiment analysis, also denominated opinion mining, which was initially focused mainly to electronic commerce but nowadays has been expanded to other fields like education, medicine, politics and others.

Due to the exponential growth of the social networks, the sentiment analysis has been strongly applied to analyze the user's opinions. Twitter is one of the social networks that has had one exponential growth in Mexico, in which is expected to have approximately 8.1 million users by the end of 2015, based on the statistics published by GlobalWebIndex. Twitter incorporates a platform of free analysis, today It has been highlighted for being useful to obtain data from the users worldwide, such data are converted to information through a process that involves natural language processing techniques, sentiment analysis, text mining and classification techniques. In Twitter the size of the comment is limited (140 characters), for that is believed those mainly could be free of spam and express a more meaningful opinion [1]. The education is one of the areas that in the last years has showed interest in analysing the comments from the students with the purpose of the professors improve their teaching techniques providing an appropriate learning to the students [2, 3]. Twitter and Facebook have been the social networks that the educational institutions have implemented to obtain feedback from their students [4].

In this paper proposes a model to analyses in comments from Twitter made by the students, who, through tweets and comments feedback teacher performance in class with the purpose of improving the teaching - learning process. One of the main tasks to do is the preparation and the process of the corpus as well as the selection of the features that finally will support the process of classification of the comments in positive, neutral and negative. This paper is organized in six sections. Section 2 describes located problem. Section 3 describes the related works. Section 4 presents our proposed Model. Section 5 presents the experimental results and the computational analysis. Finally, Section 6 presents the conclusions.

Received Nov 29, 2015 / Accepted May 21, 2016

Editorial Académica Dragón Azteca (EDITADA.ORG)

2 Problem description

Generate learning in the students is not easy for the professor because of the distractions that are around the students, in addition to the cognitive capacity of each of them, their behaviour and even the methodology which the professor implements at teaching. Most institutions of higher educational applying teacher's evaluations to know thought the students the performance of each of the professors in their academic sessions. This type of tools could allow the professors to identify their strengths, weaknesses and opportunity areas.

The Universidad Politécnica de Aguascalientes (UPA) is an institution of higher education worried for applying teacher's evaluation to improve the academic sessions and contribute the good learning to their students. The survey that the UPA applies contains a total of 21 questions in which the students evaluate through 20 closed questions about the characteristics the professor has at teaching and in the last question is permitted to insert a comment or opinion about the professor or the class (observe the figure 1).

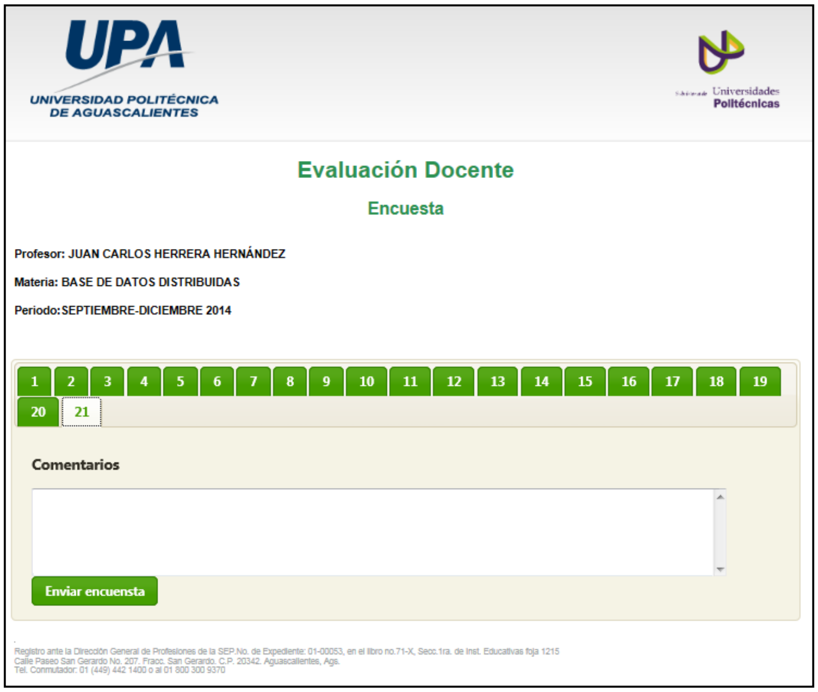

Fig. 1. GUI teacher´s evaluation

Even though this survive is applied twice in a quarter, the results obtained are a bit trustable because the majority of the students do not take the importance of the tool because they consider that their comments and recommendations are not read nor considered by the professors or the Director of Academic Program (DPA for its acronym in Spanish). The evaluation tool process consists of that every DPA delivers an Excel report that contains the results of the teacher's evaluation as well as the comments that were made by the students of every professor, however, due to there is not an informatics system that permit to analyse each comment on the students is possible that some of the comments are not perceived.

Two of the most implemented categories to represent the emotional analysis (sentiment analysis) are: document level and sentence level, which have been considered word or feature level in complementation with techniques of NLP. Numerous researchers have worked on the extraction of features with the porpoise to improve the classification process. Nevertheless, with the advantages in the field of sentiment analysis, there is no document that contains the features that must be considered for each context or application, therefore, is really important the process of selection as it is considered the base to improve the classification process, applying machine learning techniques or semantic.

3 Support Vector Machine

Support Vector learning is based on ideas which originated in statistical learning theory [5]. Most research has shown that Support Vector Machines are powerful classification tools, which can be applied to several areas. The process of how SVM transforms input to output is less clear and can be hard to interpret. For this reason, it is referred to as a black box method.

The SVM first map input data into a high dimensional feature space defined by the kernel function, and finds the optimum hyperplane that separates the training data by the maximum margin. One of the disadvantages of the SVM in solving classification problems is the global optimum, which depends on the characteristics of the dataset source. The SVM constructs a hyperplane (or a set of hyperplanes) that maximize the margin between two classes in a high dimensional space. Figure 2 shows the support vectors in the hyperplane.

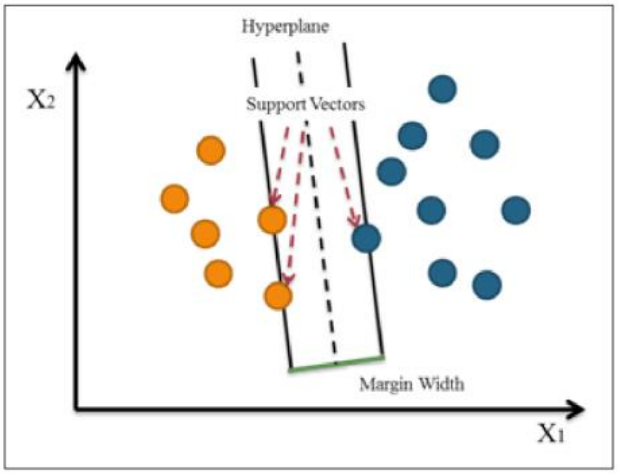

Fig. 2. Support Vector Machine [5]

The advantage of SVM is that it builds a highly accurate model through an engineering problem-oriented kernel. The two most well-known SVM tools are libsvm and SVMLite. In R we can find the implementation of *libsvm* in the e1071 [6] package and *SVMLite* in the *klaR* [7] package. Other packages are: kernlab [8] and svmpath [9].

To use an SVM is necessary to specify the kernel function, cost and gamma function. For the kernel function, the default value is radial, but it can also be specified as a linear, polynomial, radial basis, or sigmoid kernel. For the gamma argument, the default value is equal to 1/dimension, and it controls the shape of the separating hyperplane. Increasing the gamma value usually increases the number of support vectors. In case of the cost, the default value is 1, which indicates that regularization term is constant. The cost function controls training errors and margins, a large cost creates a wider margin allows for fewer misclassifications.

4 Proposed Model

In this paper is proposed to implement the SVM algorithm based on the package e1071 of the library sentiment on R created by Jurka [10], which until now includes the Naïve Bayes technique to classify emotions (fear, surprise, aversion, anger, happiness, sadness) and polarity (negative, positive, neutral). The library of sentiment includes two algorithms for the emotional classification, one of them is the Naive Bayes, which was trained by Strapparava [11] and the other consist on one simple algorithm of voter procedure. In the case of polarity classifications, the same previous algorithms are applied, however the classification algorithm of Naive Bayes for this purpose has been trained by Riloff and Wiebe [12].

In the case of the model proposed in this paper it not only intends to implement the algorithm of SVM in the sentiment library, also a corpus in Spanish labeling by polarity, weight of terms based on the tf-idf method and subjective force numerical ranging from 4-0 (4 - very good, 3 - good, 2 - neutral, 1 - negative and 0 - very bad). Comments are downloaded from the social network Twitter, where a pilot group of students have commented on the subjects identified by a hashtag. Figure 3 shows the architecture of the proposed model, in which four phases are shown, the first three phases are the first objective, the fourth phase is planned as future work to compare the results of applying machine learning classification with the results can be obtained by applying syntactic and semantic patterns.

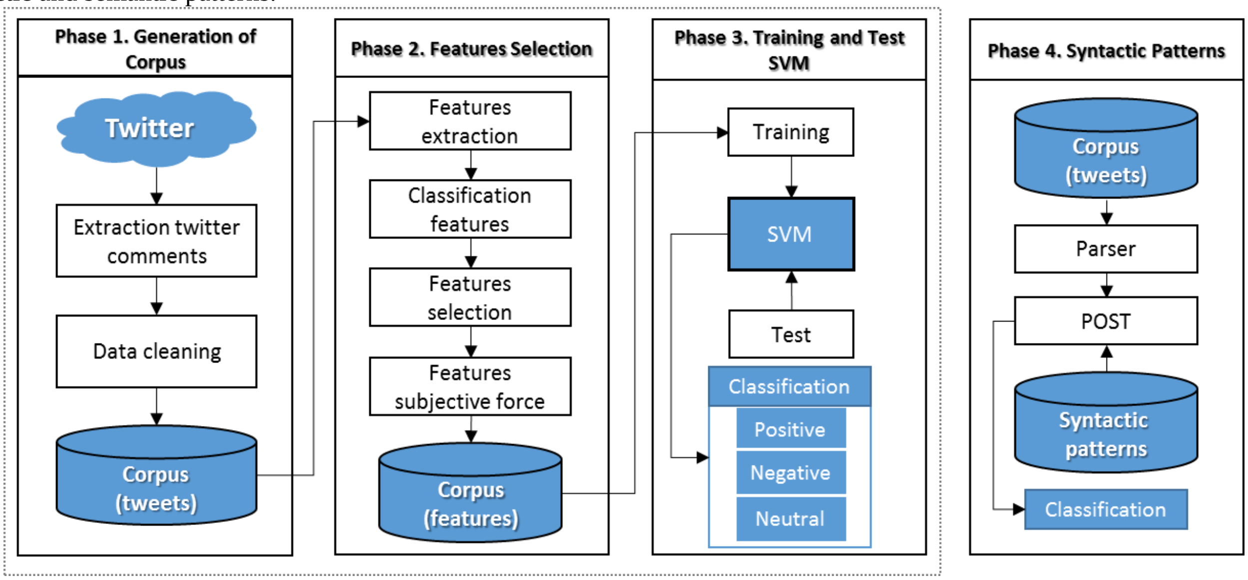

Fig. 3. Model Architecture for Sentiment Analysis Proposed

4.1 Phase 1. Corpus Generation

At level 1 of the proposed architecture tweets extraction is done by connecting the Twitter API to R program.

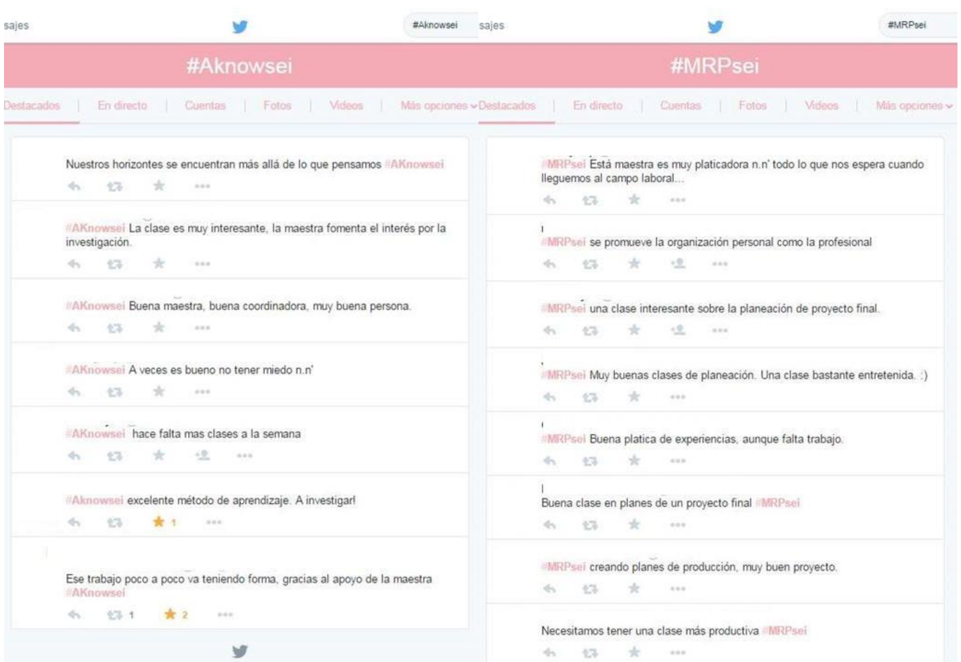

Fig. 4. Feedback tweets about teachers by the student pilot group.

Identifying tweets are carried out by a predefined hashtag to identify each subject (a label that allows to differentiate and group a specific word or topic on Twitter), which are presented in Table 1.

Table 1. Predefined hashtags for the considered subjects

No.	hashtag	No.	hashtag	No.	hashtag	No.	hashtag

1	#ASEVD01	3	#ASEVD03	5	#ASEVD05	7	#Aknowsei
2	#ASEVD02	4	#ASEVD04	6	#ASEVD06	8	#MRPsei

In the task of data cleaning the following activities are considered.

- **Tokenization Process:** the task of cutting the words of a sentence into pieces called "tokens" (instance or sequence of characters that are grouped as a useful semantic unit for processing) while removing certain characters such as punctuation.
- **Remove "stop words":** This task is responsible for removing those common words to reduce the problem of dimensionality, and to improve response time and effectiveness.
- **Remove numbers:** the numbers are removed because they most likely will not have a significant meaning in the present context.
- **Downcasing:** converting text to lowercase is done to facilitate the process of matching between words.
- **Spellcheck:** This task is also performed to facilitate the process of matching between words.
- **Process "Stemming":** this task can reduce a word to its root and is applied in order to facilitate the process of matching between words, the only detail is that this process should be verified and validated.

4.2 Phase 2. Feature Selection

At this level it takes place the process of selecting the distinctive characteristics (features), which are considered necessary extract before classification process. In the proposed model a vector of distinctive features (feature vector) based on unigram words is implemented. One of the most common methods for the selection of the features is the TF-IDF, which calculates the value of each word using the following formula:

$$wd = f_{w.d} * \log(|D| f_{W,D}) \tag{1}$$

where D is a collection of comments, w represents the terms, d the individual comments are owned by D, | D | is the size of the corpus, fw, d represents the number of times it appears w in d, FW, D is the number of comments in which w occurs in D.

Not always the most frequent words are best suited to become "features", as for example if there is a word often divided equally between classes (positive, negative) then it has no discriminative value. For this reason the application of methods or measures to reduce the error in the selection of features is needed. The features can be syntactic and semantic type which are written below.

- Syntactic feature: features such applies methods as word / POS tag, n-grams, scores, and more. In his research Fei [13] noted certain patterns in phrases to detect the polarity of these, since it states that the case of $n + j$ (noun followed by a positive adjective) probably the phrase has a positive orientation, however itself contains n-dj (noun followed by negative adjective), then the phrase expresses a negative sense.
- Semantic feature: the semantic features have a score provided based on their lexical meaning leaning dictionaries as WordNet.

The process of extracting features Selvam [14] indicates that you can deal with some problems like: type of feature, selecting features, the weight and reduction features, which are described below.

- Type feature: 1.- TF (give a weight depending on their frequency of appearance in the commentary or document), 2.- Term Co-occurrence (unigrams, bigrams or n-grams) 3.- POST (label agreement to the corresponding grammar) 4.- Emotion (label according to polar emotion expressed), 5. Denial (terms that can change the direction of the phrase) and 6. - Syntactic Unit (represented by syntax tree).
- Select feature: the selection of appropriate features reduces the problem of dimensionality. This can be done using any of the following methods: 1. - Gain information (based on the presence and absence of terms in a limit of comments, those terms considered irrelevant for their informational value are eliminated), 2. - DF (measures the number of occurrences of a term in the available amount of comments). 3. - Mutual information (selects only those terms with the frequent association in the comments).

- Weighting of features: two methods of associating a weight to the features are: the presence and frequency of the term (PFT, which indicates that there may be less frequent terms that have potential information), TF-IDF (method by which weighted higher value of words that appear mostly in the comments).
- Reduction features: reducing features allow a better performance in qualifying.

After the selection of the features is necessary to perform a Term-Document Matrix, which represents the relationship between the terms and comments, each row contains a term and each column contains the comment and as input is the number of occurrences of each term in a commentary. After construction of the Term-Document Matrix can analyze the importance of the terms by generating a wordcloud, calculating the frequency of words. This method also allows to refine the features for debugging the classification process.

4.3 Phase 3. Training and Test SVM

For the classification by SVM technique it is performed a supervised learning whit two corpus: training set and test set. For satisfactory results with SVM is important to select the features which contain the most distinctive properties of the text and then convert them to numbers. Each feature has a label indicating polarity and a weight that indicates its intensity, for example: *"the kind of quality software is vital to our*specialty,"in this sentence the word *"vital"* would be the feature with a greater weight to show a positive polarity. The weights which may possess features ranging from 4 to 0, where 4 shows a positive value, the 3 is positive, 2 neutral, 1 is negative and 0 is very negative.

Given a set of opinions labelled in certain subjects A_1 ... A_n., One of the objectives of the model is to classify subjects A_{n+1} based on comments labeled. The labeling corpus of a subject comments i is represented by D_i. Where D_i contains two elements (r_{ij}, l_{ij}), r_j is the comment and l_j has the tag comment (4-0).
To evaluate the performance classification of SVM it is using the following equation:

$$Efficiency = \frac{(Real\ positives + Real\ negatives) * 100}{Total\ comments} \quad (2)$$

in the above equation efficiency is the radius of the sum of the real positives and real negatives between the total comments from the test dataset.

4.4 Phase 4. Syntactic Patterns

Phase 4 called Syntactic Patterns is proposed as future work to enable comparison between the results obtained with the results using SVM and syntactic patterns. For this phase applies part-of-speech tagging, POS tagging or POST to the corpus generated in phase 1, using for this task library Stanford POS tagger[1].,where they have to select adjectives, adverbs and verbs like features to detect emotions in the comments. About this Rahate [15] indicates that adverbs denote semantic orientations and adjectives and verbs often represent emotions. To determine the value (score) of adjectives, adverbs and verbs extracted it will be used a lexicon or an affective dictionary preferably available in Spanish [16, 17].

The POST process is done in Java and is looking for the labeling of each term of the sentence in accordance with EAGLES standard, in order to automatically identify adjectives, adverbs and verbs and their position in the sentence. The following table shows an extract of the labels to be considered and their respective description.

Table 2. Tags used by EAGLES Standard

Code label	Description	Example
ao0000	adjective (ordinal)	
aq0000	adjective (descriptive)	
Rg	adverb (general)	siempre, más, personalmente
Rn	adverb (denier)	no
vmn0000	verb (infinitive)	dar, trabjar

Because verbs that make up a sentence can be conjugated at different times, there’s a variety of codes labels proposed by EAGLES standard to identify them, for example: *vag0000* belongs to those auxiliary verbs gerund, *vam0000* corresponds to the auxiliaries imperatives and so on; for example the phrase *" es un profesor muy bueno para exponer temas teóricos y te enseña muy bien en lo práctico."* has the following results in the process of tagging.

```
es/vsip000 un/di0000 profesor/nc0s000 muy/rg bueno/aq0000 para/sp000 exponer/v
temas/nc0p000 teóricos/aq0000 y/cc te/pp000000 enseña/vmic000 muy/rg bien/rg en
lo/da0000 práctico/nc0s000 ./fp
```

Then the result of parsing the same phrase is shown *("es un profesor muy bueno para exponer temas teóricos y te enseña muy bien en lo práctico.*").

```
(ROOT
  (sentence
    (S
      (grup.verb (vsip000 es))
      (sn
        (spec (di0000 un))
        (grup.nom (nc0s000 profesor)
          (s.a
            (spec (rg muy))
            (grup.a (aq0000 bueno)))))
      (sp
        (prep (sp000 para))
        (S
          (infinitiu (vmn0000 exponer))
          (sn
            (grup.nom (nc0p000 temas)
              (s.a
                (grup.a (aq0000 teóricos))))))))
    (conj (cc y))
    (S
      (sn
        (grup.nom (pp000000 te)))
      (grup.verb (vmic000 enseña))
      (sadv
        (spec (rg muy))
        (grup.adv (rg bien)))
      (sp
        (prep (sp000 en))
        (sn
          (spec (da0000 lo))
          (grup.nom (nc0s000 práctico)))))
    (fp .)))
```

This future phase to implement is addressed to linguistic analysis based on established patterns to identify the polarity in a sentence according to the dependency that exists between the terms, considering the grammar and semantics [13, 18].

5 Experimental studies

Comments downloaded from Twitter are corresponding to 8 subjects imparted to the pilot group of Engineering in Strategic Information Systems of UPA, the extension of the comments is maximum 140 characters because of the Twitter rules. In the first instance the stop words are removed since they do not significantly influence how the words relate to each other, can be prepositions, pronouns, articles, adverbs, conjunctions and some verbs. Figure 6 shows some stop words that are removed in the comments corpus. It should be noted that the search engines like google has ignored such terms when indexing a website.

a	conseguir	emplean	fin	ir	podría	sabemos	tal
acá	consigo	emplear	fue	junto	podríais	saber	también
ahí	consigue	empleas	fueron	la	podríamos	se	tan
ajena	consiguen	empleo	fui	los	podrían	según	tanta
al	consigues	en	fuimos	me	por	ser	te
algo	contigo	encima	ha	mi	por qué	sí	tenéis
algún	cual	entonces	hacéis	mía	porque	siendo	tenemos
allá	cuales	entre	hacemos	mientras	primero	sin	tener
ambos	cualquier	eramos	hacen	mío	puede	sino	tengo
ante	cuando	eran	hacer	misma	puedo	so	ti
antes	cuanto	eres	hacia	modo	pues	sobre	tiempo
aquel	de	es	hago	nos	que	sois	tienen
aquella	dejar	esa	hasta	nosotras	qué	sólo	toda
aquí	del	esta	incluso	nuestra	querer	somos	tomar
arriba	dentro	estaba	intenta	nunca	quién	soy	trabaja

Fig. 5. Example of list of stop words to remove

5.1 Corpus of features (adaptation)

The corpus considered as a basis for adapting the features corpus to be implemented in the model is Riloff and Wiebe [12], which is available at the sentiment library (subjectivity.csv) and consists of three columns, where it is the term, the subjective weight of the term (weaksubj, strongsubj) and the polarity (positive, negative).

Table 2. Wiebe's corpus [12]

Term	Subjective Force	Polarity
abandonment	weaksubj	negative
abandon	weaksubj	negative
abase	strongsubj	negative
abasement	strongsubj	negative
abash	strongsubj	negative
abate	weaksubj	negative
abdicate	weaksubj	negative
aberration	strongsubj	negative
abhor	strongsubj	negative

Because analyze tweets about teacher evaluations involves words or text relating to the context of education, distinctive terms are added to support the classification process called features. This is because the original corpus contains very general terms and not focused on education. For example Table 3 presents 10 tweets about teacher evaluations.

Table 3. Fragment of tweets about teacher evaluation

No.	Tweets	Class
1	realiza *muy* **entretenida** su clase	Positive
2	un *buen* maestro, sin **duda** alguna tiene un **dominio** de la clase y del **conocimiento** del tema	Positive
3	la clase es *muy* **completa**	Positive
4	*buen* profesor y *bien* **preparado**	Positive
5	muy **buena** clase, el profesor es *muy* **bueno** y **sabe resolver** las **duda**s	Positive
6	es una materia *muy* **útil** y *muy* **interesante**	Positive
7	**no respeta** el reglamento	Negative
8	un profesor **malo** en realidad	Negative
9	**no** tiene total **dominio** de los temas	Negative
10	**no** he **aprendido** nada y además es *bastante* **aburrida** la clase	Negative

In Table 3 the potential terms to become features are highlighted in bold and the quantifiers terms are represented in italics. For each context it is important to identify the stop words due to the removal actions applied by some search engines (eg. Google, Ask) or their implementation in different text recovery systems. When using a standard list available in online runs there is a conjoint risk that searches or in this case classifications are ineffective and hence magnify computational complexity. When using the model the stop words are also removed before selecting features so they do not stand out in the word cloud process neither in the frequency terms table. Similarly before selecting features the quantifiers and some substantive are removed. Figure 6 shows the word cloud made by a corpus of about 800 Twitter comments by students about their classes. The word cloud (a) presents the most frequent terms of the corpus, the word cloud (b) also shows the most frequent terms of the corpus; however it does not consider the stop words, quantifiers and some nouns.

a b

Fig. 6. Word cloud (a) with stop words includes (b) without stop words

Once the word cloud is done, we proceed to make an accounting of the common terms.

- To calculate TF the following formula is used.
- TF (term) = frequency of the term in the document (comment) / no. of terms in the document (comment).
- To calculate IDF the following formula is used: IDF (term) = log (no. of comments in the corpus / comment frequency of the term).
- To calculate TF-IDF of each term the following formula is used: TF-IDF (term) = TF (term) X IDF (term).

For example Table 4 displays a fragment of calculating terms frequency and TF -IDF in a corpus of about 500 comments, where the term *excelente* has a higher frequency and TF -IDF weight.

Table 4. Results of the TF-IDF process

Term	TF(weight)	TF	IDF(weight)	TF-IDF(total)
excelente	0.1420765	77	3.8338758	0.54470367
falta	0.06739526	37	4.89120897	0.32964432
sabe	0.0564663	30	5.19377174	0.29327309
explica	0.05464481	29	5.24268134	0.28648532
dudas	0.05282332	29	5.24268134	0.27693581
dominio	0.04918033	27	5.34577484	0.26290696
explicar	0.03642987	20	5.77873424	0.21051855
conocimiento	0.03460838	19	5.85273483	0.20255366
dar	0.03460838	19	5.85273483	0.20255366

más	0.03278689	17	6.0131995	0.19715408
clara	0.03096539	17	6.0131995	0.18620108
aprendizaje	0.0291439	16	6.10066234	0.17779708
dinámica	0.0291439	16	6.10066234	0.17779708
gusta	0.0291439	16	6.10066234	0.17779708

To embody the features corpus the following attributes must be considered: term or feature , subjective force , polarity, tf- idf and numerical subjective force, which are described below.

- **Term or feature**, In the first instance all terms of the corpus created by Riloff and Wiebe [12] were translated, corpus implemented in the sentiment R library, and those potential terms to become features were added for the classifications tweets of teacher evaluation (see Figure 7).

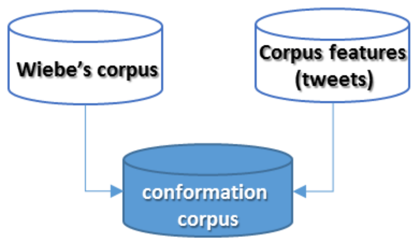

Fig. 7. Conformation of corpus with features

- **Subjective force**, It indicates the subjective force of the term to assist in the classification process.
- **Polarity,** It indicates whether the term is positive, negative or neutral.
- **Td-idf**, It indicates how relevant the term in the corpus of Twitter comments is.
- **Numerical subjective force**, It indicates the subjective force of the term assigned to assist in the classification process with the difference that it can take a value between 4 to 0, where 4 is given to very positive terms, 3 is for positive terms, 2 for neutral terms, 1 to negative terms and 0 for very negative terms; this is implemented in order to improve the classification process due to it could be positive terms identified as subjective strongly, as well as negative terms in the same case, this can lead to a confusion in the classification, so it was decided to add a numerical subjective force. Table 5 presents a fragment of the modified corpus liable for implementation.

Table 5. A fragment of the corpus with features

TERM	SF	Polarity	N-SF	TF-IDF
Ayuda	strongsubj	positive	3	0.15155354
atento	strongsubj	positive	4	0.04653388
dedicado	strongsubj	positive	4	0.0331536
excelente	strongsubj	positive	4	0.54470367
aburridas	strongsubj	negative	0	0.10347049
entretenida	strongsubj	positive	4	0.0331536
fomenta	strongsubj	positive	4	0.11361865
interés	strongsubj	positive	4	0.01839829
opinar	weaksubj	positive	3	0.0331536
estricto	weaksubj	neutro	2	0.0331536
bueno	strongsubj	positive	4	0.16048507
debería	weaksubj	negative	1	0.14241967
enseña	strongsubj	positive	4	0.12347421
mal	strongsubj	negative	0	0.11361865

6 Results and Conclusions

A corpus of commentaries on teaching evaluations by a pilot group of an academic period comprised from May - August 2015 with a total of 2000 comments was used for testing of the classification process, that corpus was labeled manually by two evaluators and the following results were obtained: 1000 comments negative comments, 10 neutral and 990 positive comments. Figure 8 displays the results when evaluating 2000 comments on the corpus named subjectivity, originally created by Riloff and Wiebe [12] and implemented in Jurka's library [10]. The subjectivity corpus was translated into Spanish and adapted to the context of this research.

Figure 8 represents the first result sorting the corpus of comments, it can be seen that over 600 comments are classified as neutral. As a previous step on the SVM implementation the subjectivity corpus has to be adapted; therefore, the tests were conducted using Naive Bayes classifier, which is the method originally considered in the Sentiment library.

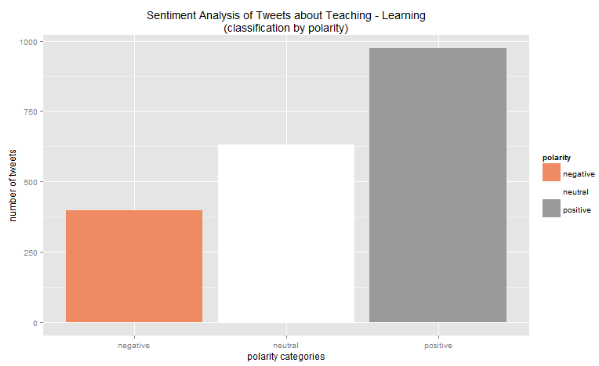

Fig. 8. Results of labelling comments

Due to the first results obtained and shown in Figure 8 it is considered necessary to make a review of the subjectivity corpus and remove repeated words, as well as words with more than two terms and add the most representative features of teacher evaluation comments. The following results were obtained when applying the previous process (see Figure 9).

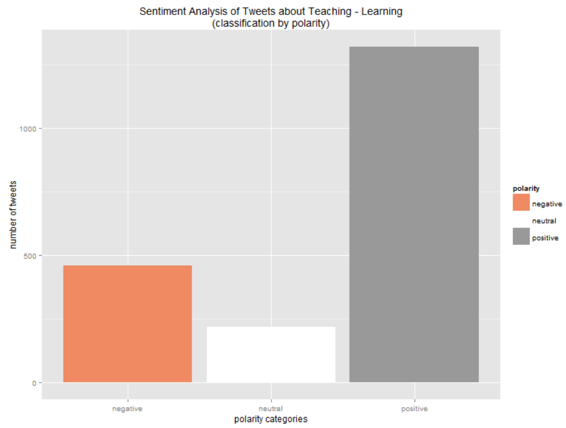

Fig. 9. Results of labelling comments v2

The most common cause for mislabeling is that subjectivity corpus doesn't have all the terms that somehow influence the classification; however addition of each of them besides its gender: masculine, feminine and plural, for example *aburrido*, *aburrida*, *aburridos*, is quite a time consuming task, therefore the experiment of adding the most influential terms was conducted for the classification and then the stemming process was performed to compare the results of the classification (see figures 10 and 11).

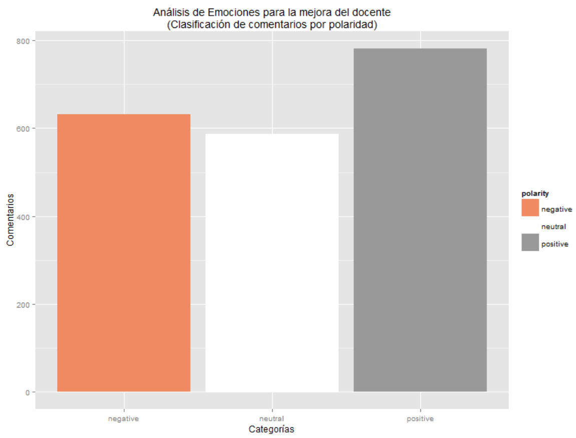

Fig. 10. Results with influential terms.

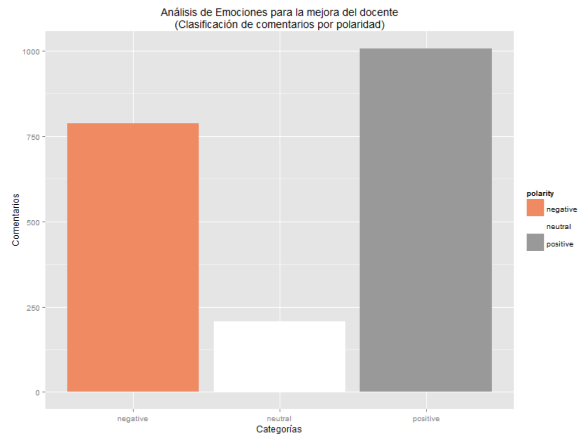

Fig. 11. Results with stemming process in subjectivity corpus and in the comments.

The stemming process improved considerably the classification of teacher evaluation comments because by reducing a word to its root or a stem, the matching process worked better, but as can be seen in Figure 11 it is still necessary to perform a verification and validation process in the results, in order to identify the lack of reliability displayed. Once the training process is complete with the subjectivity corpus focused on the context of teaching evaluations, this will be implemented in the SVM model, in order to measure the efficiency between the two techniques, as well as other machine learning techniques to apply the most effective.

The adequacy of the subjectivity corpus and the choice of the classification technique will contribute to the creation of a Model that support to universities to provide high quality and vanguard education considering students' opinions about classes and their teachers. The evaluation of comments allows teachers, based on feedback from students, to improve their classes. There is also an observed potential in the implementation of courses to improve teaching.

References

1. Junco, R., Heiberger, G., & Loken, E., *The effect of Twitter on college student engagement and grades.* . Journal of computer assisted learning. Volume 27, Issue 2, pages 119–132, April 2011, 2011.
2. Altrabsheh, N., Gaber, M. and Cocea, Mihaela, *SA-E: Sentiment Analysis for Education.* In: 5th KES International Conference on Intelligent Decision Technologies, 2013-06-26 - 2013-06-28, Sesimbra., 2013.
3. Turney, P., *Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews.* ACL '02 Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Pages 417-424 2002.
4. Nan, L.a.D., D., *Using text mining and sentiment analysis for online forums hotspot detection and forecast.* Decision Support Systems archive. Volume 48 Issue 2, January, 2010. Pages 354-368. Elsevier Science Publishers B. V. Amsterdam, The Netherlands, The Netherlands, 2010.
5. Vapnick, V., *Statistical Learning Theory*. Wiley, New York, 1998.
6. Dimitriadou E, H.K., Leisch F, Meyer D, Weingessel A, *e1071: Misc Functions of the Department of Statistics.* TU Wien, Version 1.5-11." URL http://CRAN.R-project.org/. 2005.
7. Roever C, R.N., Luebke K, Ligges U, *"klaR – Classification and Visualization.".* Version 0.4-1. URL http://CRAN.R-project.org/. 2005.
8. Hsu CW, L.C., *A Simple Decomposition Method for Support Vector Machines.* Machine Learning, 46, 291–314. URL http://www.csie.ntu.edu.tw/~cjlin/papers/decomp.ps.gz., 2002.
9. Hastie, T., *svmpath: The SVM Path algorithm.* R package, Version 0.9. URL http://CRAN.R-project.org/. 2004.
10. Jurka, T., *Sentiment: Tools for Sentiment Analysis.* URL http://CRAN.R-project.org/package=sentiment. R package version 0.1., 2012.

11. Strapparava, C., Valitutti, A. and Stock, O., *The Affective Weight of Lexicon.* In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006), Genoa, Italy (May 2006), 2006.
12. Riloff, E.W., J., *Learning extraction patterns for subjective expressions.* In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, EMNLP 2003, pp.105-112., 2003.
13. Fei, Z., Liu, J., Wu, G., *Sentiment classification using phrase patterns.* Computer Society, Los Alamitos, pp. 1147-1152. IEEE, 2004.
14. Selvam, B.a.A., S., *A Survey on Opinion Mining Framework.* International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013, 2013.
15. Rohini S. Rahate , E.M., *Feature Selection for Sentiment Analysis by using SVM.* International Journal of Computer Applications 12/2013; 84(5):24-32. DOI: 10.5120/14573-2697, 2013.
16. Esuli, A.S., F., *SENTIWORDNET: a publicly available lexical resource for opinion mining.* . In Proceedings of the 5th Conference on Languaje Resources and Evaluation (LREC 2006). pages 417-422, Genoa, Italy., 2006.
17. Carrillo de Albornoz, J., Plaza, L. & Gervás, P., *SentiSense: An easily scalable concept-based affective lexicon for Sentiment Analysis.* The 8th International Conference on Language Resources and Evaluation (LREC 2012), 2012.
18. Kien-Weng, L., Na, J., Theng, Y. and Chang, K., *Phrase-Level Sentiment Polarity Classification Using Rule-Based Typed Dependencies and Additional Complex Phrases Consideration.* Tan LKW, Na JC, Theng YL et al. Phrase-level sentiment polarity classification using rule-based typed dependencies and additional complex phrases consideration. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 27(3): 650–666 May 2012. DOI 10.1007/s11390-012-1251-y, 2012.