Bilingual and Cross Domain Politics Analysis

Jean-Valère Cossu[1], Rocío Abascal-Mena[2], Alejandro Molina[4],
Juan-Manuel Torres-Moreno[1,3] and Eric SanJuan[1]

[1]LIA/Université d'Avignon et des Pays de Vaucluse, Avignon,
France
[2]Universidad Autónoma Metropolitana-Cuajimalpa, México DF,
México
[3]École Poytechnique de Montréal, Montréal,
(Québec) Canada
[4]Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, México DF,
México
rocio.abascal@gmail.com, molina@conabio.gob.mx,
{jean-valere.cossu, juan-manuel.torres, eric.sanjuan}@univ-avignon.fr

Abstract. Opinion mining on Twitter recently attracted research interest in politics using Information Retrieval (IR) and Natural Language Processing (NLP). However, getting domain-specific annotated data still remains a costly manual step. In addition, the amount and quality of these annotation may be critical regarding the performance of machine learning (ML) based systems. An alternative solution is to use cross-language and cross-domain sets to simulate training data. This paper describe a ML approach to automatically annotate Spanish tweets dealing with the online-reputation of politicians. Our main finding is that a simple statistical NLP classifier without in-domain training can provide as reliable annotation as humans annotators and outperform more specific resources such as lexicon or in-domain data.

1 Introduction

Sentiment Analysis (SA) is useful in the study of online communication because it gives researchers the ability to automatically measure emotion in online texts [30]. The rise of social web, particularly Twitter[1], provides new opportunities to collect real time data in large quantities directly from users. Tweets can be analysed in order to track reactions to events. One important aspect of the tweets is that they have been used as a way to participate in social movements as well to make public opinions and reactions to different events. These opinions are charged with sentiment whether they can be positive or negative, toward a movement or event, as they can change over time.

Since Twitter provides the possibility to extract tweets and compose actual corpus there have been a lot of linguistic research applied in tweets. Using publicly available online data to perform sentiment studies significantly reduces

[1] http://www.twitter.com

the costs, efforts and time needed to administer large-scale public surveys and questionnaires [8]. Specially, we find that Twitter is often used to analyse political preferences by studying the use of humor contained in tweets [8] where POMS (Profile of Mood States) was used to distill, from Twitter messages, time series corresponding to 6 different emotional attributes (tension, depression, anger, vigor, fatigue and confusion). POMS is a psychometric instrument that provides a list of adjectives for which the patient has to indicate the level of approval. Each adjective is related to a state of mind and, therefore, the list can be exploited as the basis for a mood-analyser of textual data.

Politics have already been addressed in previous works but mostly in English as [22] or [35]. [23], used a subjective lexicon that comes from the Opinion Finder in order to determine positive and negative scores for each data set corresponding to a tweet. In this case, the raw numbers of positive and negative tweets about a given topic are used to calculate a confidence score (the relation between the number of positive and negative tweets). The authors indicated that by a simple manual inspection of the tweets they have found examples that have been classified incorrectly based on a feeling. Nevertheless, the authors used this method to measure some issues such as the consumer confidence, the presidential approval and the 2008 presidential election in the United States.

To the extend of our knowledge, political studies from a ML point-of-view in Spanish or French are rare [34]. However the role of social networks during the Presidential campaign of 2012 in Mexico gained great importance as the principal instrument for exercising public opinion, especially for young people. The youth organization "yosoy132" born during the election campaign in Mexico in 2012, thanks to the social networks, joining youth groups from all universities in the country regardless their social conditions. Studies conducted by the *Instituto Nacional de Estadística, Geografía e Informática* (INEGI)[2] show that 40.3% of users of ICT in Mexico are young people who communicate via social networks and mobile phones and that they remain connected most of the day. This percentage equals 15.3 million people aged 18 to 34 who were potential voters of the political election [29]. In this way, in 2012, we found that participation via Twitter increased creating significant social and political communities around specific problems.

Usual studies in the domain assume that a great effort of acquisition of the tweets and a subsequent manual labeling process is required. In addition, a validation process is needed to correct the errors introduced by manual labeling. Even using crowd-sourcing-based solutions it is a very expensive task both in money and time. Moreover important political events will always occur faster than our capacities of getting manually annotated data in several languages. In this context, we propose an approach that can provide a reliable pre-annotation using out-of-domain data which needs only a light supervision before validation in order to obtain a reliable corpus that can be used for more complex political studies such as user Political Tendency or to monitor politicians reputation.

The rest of this paper is organized as following. Section 2 gives a focused

[2] National Institute of Statistics, Geography and Informatics of Mexico.

overview of related work. In section 3 we describe the main characteristics of the data-set. In section 4 we propose our approaches while section 5 is devoted to a thoroughly evaluation. Finally, section 6 gives some conclusions about our work and opens several perspectives.

2 Tweets mining and sentiment analysis

[32] presented in 2010 a job with two distinct parts: in the first one LIWC (Linguistic Inquiry and Word Count) is used to perform a superficial analysis of the tweets related to the different political parties that competed for the German Federal election in 2009. In the second part, the authors claim that the mere counting of tweets with references to one of the parties, accurately reflects the election results. On the other hand, they established that the MAE (Mean Absolute Error) of the "prediction" based on Twitter data was very close to the real surveys that were carried out.

An increasing number of empirical analyses of sentiment and mood are based on textual collections of data generated on Twitter as they used sophisticated algorithms to pre-process, apply grammatical rules and classify them in mood categories. In this way we find, for example, the use of a lexicon based classifier as a dataset is also classified using SVM and/or Naive Bayes [36,20]. A classifier is developed specifically tuned for tweets, using key words, phrases and emoticons to determine the mood of each tweet [6]. Several methods have been already proposed for exploiting tweets in order to detect people's mood changes throughout the day [20,19].

[10] has measured changes in the mood of the U.S. population, over three years, from tweets, providing policy relevant indicators. In general, studies analysing tweets by combining different sentiment analysis algorithms have been able to give new insights into human behaviour as a result [11,12,17,18]. These works show that there is a tremendous ambition to develop opinion mining tools for social media in order to be able to distinguish what is important and interesting [21].

The research field of SA, also known as opinion mining, has developed many algorithms to identify whether an online text is subjective or objective, and whether any opinion expressed [24]. Another way to identify polarity is based on the use of lexicons. There are lexicons like SentiWordNet[3] [3] that attaches real-valued sentiment scores to WordNet synsets [15]. [7] maintains and freely distributes a sentiment lexicon which contains morphological variants, slang, and social-media mark-up in order to be able to analyse sentences even if they are misspelling.

Nevertheless theses approaches miss an important aspect of Spanish language and politics domain: the irony. Irony is the creative use of language in order to make fun of something or someone. The boundaries between irony and sarcasm are vague. It seems that Spanish is a language that uses a lot of irony in the

[3] `http://sentiwordnet.isti.cnr.it/`

communications [1]. The Spanish language considered irony a "soft and veiled mockery", while sarcasm is a "hard and bloody mockery". Thus, the only difference is the degree of cruelty associated. For example, to ironize over a retarded person saying: "He is a genius" would be a sarcasm, and the same remark referred to a colleague who has been successful by chance solve a problem, an irony. Sarcasm is then an extreme form of irony [1]. Detection of irony is a difficult subject, and has been subject matter of various disciplines such as linguistics and psychology, among other [26]. Studies of [5] and [4] on Tweeter data shown ironic English tweets classification using Decision Trees. Sarcasm detection in tweets and Amazon has been realized by SASI [13] using the Mechanical Turk to create a gold-standard. To our knowledge, there are no studies of irony or sarcasm on Spanish or French Tweets.

While [9], [33] and [34] obtained agreement percentage quite similar to other studies over sentiment analysis task, they all agreed that human interpretation of these kind of more or less consensual contents is prone to mistakes. As both facts and opinions have to be considered, regardless of whether the content is opinionated or not, it is sometimes hard to tell what implications a message may have on the reputation of a given entity. The political context finally makes the task harder.

In this work we investigate how much ML without correct training data can perform compared to humans annotators.

3 Approaches

We mainly used two approaches in this paper: the lexicon approach and the Machine Learning approach.

3.1 Lexicon approach

Our collection was analysed by using a lexicon approach combined with a linguistic analysis in order to detect sentiments, during a period of time, in social and political tweets. The lexicon approach starts with a list of positive and negative words, which are already pre-coded for polarity. A linguistic analysis, in contrast, exploits the grammatical structure of text to predict its polarity, in conjunction with the lexicon [30]. Words contained in a tweet are classified into positive or negative by using the previous lexicon. Nevertheless, this methodology does not takes into account the sarcasm which transforms the polarity of an apparently positive or negative utterance into its opposite [16]. But by analysing a big corpus the sarcasm rest minimum and do not contributes to inflate in a big amount the percentage of the total results. The corpus is pre-processed in order to extract stop-words, punctuation, links, etc. Then, the Spanish and the English translated lexicon, respectively, are used to count for each tweet and for each corpus the number of positive and negative words contained in each tweet. All the process is automatically performed by using R[4].

[4] R is an interpreted computer language designed for statistical data analysis (`http://www.r-project.org/`

3.2 Machine learning approaches

As the process described above is not sufficient, we propose an alternative approach using probabilistic ML. In works such as [9] and during TASS [34] ML was partly used to assist annotators and propose annotations. [14] showed that a small annotated set coupled to ML could perform competitively to annotators to answer text mining tasks. The annotation was addressed as a classification problem that consisted of determining the polarity of each tweet. A very large collection of ML algorithms can be used for classification tasks. We chose to use a multi-class SVM-based approach[5], a Cosine similarity and a baseline. The baseline was computed as simple memory test which consists in tagging each tweet with the most similar tweet in the training set (according to Jaccard index). All these approaches rely on a discriminant bag-of-words representation (DF, IDF and Gini) [31] of each tweet (we considered n-gram with $n \leq 3$). The bag-of-words is built after the following pre-processing:

– words are lower-cased;
– stop-words[6], links and punctuation are removed;

Then we estimate the similarity of a given tweet by comparing it to each class's bag-of-words and rank tweets according to the classifier value.

4 Datasets and Descriptive Statistics

In this section we provide a short description of each data-set we used. To investigate to classification issue we wanted to compare the use of similar data with a French annotated corpora of tweets dealing with politics and with a collection of tweets dealing with corporate entities reputation.

4.1 Description of the Spanish political set

The corpus analysed concerns 800 tweets containing #AMLO that were extracted between the period of 9, 10 and 11 June 2012. AMLO is the acronym for Andrés Manuel López Obrador, who was a left candidate to the Presidential elections in Mexico. AMLO has built a strong base of support among people who feel that they have been left behind as Mexico's economy grows and evolves. He had the loyalty of a young generation who were frustrated with the country's monopolistic media sector (mainly represented by Televisa[7]).

As shown in table 1 classes are well balanced with only a slightly difference with negative tweets for the complete collection as well as for the French subpart.

[5] Multi-Class Support Vector Machine, see: `http://www.cs.cornell.edu/people/tj/svm_light/svm_multiclass.html`
[6] Stop-list from Oracle (`http://docs.oracle.com`).
[7] Which currently dominates approximately 70% of the Mexican broadcasting and pay-TV market: `http://www.nasdaq.com/article/strict-telecom-laws-to-check-mexican-duopoly-analyst-blog-cm334020`

Table 1. Class distribution in both complete and French sub-part collection

Class	Class-Distribution	Class-Distribution (French)
Negative	0.41	0.37
Neutral	0.29	0.30
Positive	0.30	0.33

4.2 Statistics on the French political

We provide here some details about the French data-set (more detailed statistics are available in [33]) about the two main candidates at the last french presidential election (May, 2012). Tweets were extracted from March 2012 (before the election) to December 2012 (after the election) This collection has been manually annotated by thirty post-graduate persons regarding polarity (6 levels) and topic. It comprises more than 11,527 annotations (7,283 unique tweets, half have been annotated twice or more). In our study we only consider the polarity annotation and for a reasonable analysis, we considered only 3 polarity level.

Table 2. Class distribution in the French political collection

Class	Class-Distribution
Negative	0.60
Neutral	0.12
Positive	0.28

Table 2 shows that the main tendency is negative wit a very few number of neutral tweets. The main reason is that politics in France unleash passions between people.

4.3 Statistics on the Spanish annotated set

We used the Spanish sub-part (23,100 tweets of the RepLab 2013 campaign [2]) collection which is a large multilingual collection of tweets referring to a set of 61 entities spread into four domains: automotive, banking, universities and music/artists. For each entity, at least 2,200 tweets have been collected covering a period going from the 1^{st} of June 2012 to the 31^{st} of December 2012. These tweets have been manually annotated by experts according to filtering (related or not to the entity), polarity for reputation (3 levels), topic (clustering) and priority (topic ranking). We will only use the polarity annotation. Table 3 shows that the main tendency of the RepLab set is positive. Crossing this point with the negative view from the French collection should provide an interesting result.

Table 3. Class distribution in the Spanish reputation collection

Class	Class-Distribution
Negative	0.24
Neutral	0.28
Positive	0.48

4.4 Metrics

The measures selected to evaluate our approach were the Precision, the Recall, and the F-Score for each class. The F-Score reported in tables 4 and 5 is the the Macro Averaged F-Score computed as mean of each class F-Score.

$$\text{F-Score}_c = \frac{2 \times (\text{Precision}_c \times \text{Recall}_c)}{\text{Precision}_c + \text{Recall}_c} \tag{1}$$

With Precision P_c for class c as:

$$P_c = \frac{\text{Num. of documents correctly assigned in class}_c}{\text{Num. of documents assigned in class}_c} \tag{2}$$

And Recall R_c for class c as:

$$R_c = \frac{\text{Num. of documents correctly assigned in class}_c}{\text{Num. of documents belonging to class}_c} \tag{3}$$

Results are also compared according to Accuracy as it is a easy interpretable measure and it represents the informativeness of a system.

5 Experimental evaluation

Tables 4 and 5 summarize the experimental results of our proposal concerning the tweets polarity.

5.1 Classification using translated data

Table 4. Polarity classification results using French set

Method	F-Score	Accuracy
Baseline	**0.39**	**0.42**
Cosine	0.24	0.36
SVM	0.33	0.37

As a first experiment we choose to manually translate a sample (220 tweets) of our unlabeled set in order to perform a classification using the French annotated set training set. According to Table 4 systems performance using same context data is really low. The main reasons are that the vocabulary used to described the French candidates is not the same as the one used for AMLO but also that both class distributions are too different.

5.2 Classification over Spanish contents

Table 5. Polarity classification results using Spanish set

Method	F-Score	Accuracy
Baseline	0.50	0.51
Cosine	**0.74**	**0.74**
ElhPolar Lexicon	0.25	0.32
Translated Lexicon	0.21	0.33
SVM	0.17	0.31

In a second experiment we performed the classification using RepLab 2013 reputation set. Table 5 shows classification performance over Spanish contents according to F-Score and Accuracy. An interesting result is the systems' ranking while the Cosine similarity was outperformed with the French sets it is interesting to notice that we are able to obtain quite good classification results that are close to inter-annotator agreements observed in the literature [2,34,25], while SVM performance dramatically decreases. Baseline performance is also quite interesting since his performance increases but remains lower than the Cosine it was significantly better on the French set. Both lexicon approaches (ElhPolar lexicon [27] and Bing Liu translated one) seem to not fit our data-set or this kind of analysis since they do not perform well.

Some contents such as:
"RT 1. Naces 2. Eres AMLO 3. Creces 4. No eres presidente. 5. No eres presidente. 6. No eres presidente. 7. No eres presidente. 8. Mueres. JAJA" (In English: 1. You're born 2.You're AMLO 3. You grow 4. You're not president. 4. You're not president. 6. You're not president. 7. You're not president 8. You die. LOL LOL")
are tagged positive by the systems while they are really negative for AMLO. It is a typical example of humoristic contents that systems are not able to handle properly.

In this another example:
"AMLO gran orador cada vez que abre la boca sueña #elpejeaburrehastaalospejezombies" ("AMLO great speaker every-time he opens the mouth he dreams" in English)
is an irony because people are not dreaming about a better country instead they are becoming tired and almost falling at sleep every time that AMLO speaks.

6 Conclusions

In this paper we described and compared several approaches for a fast political classification of Spanish tweets. Our experimental evaluation (although our test set was limited) establish that without specific training material we can achieve results comparable to state of art. Then, while the literature insists on the need of specific training data, our experiments have shown that the need of costly experts annotation can be reconsidered.

At first we intend to apply this process to others candidates and other existing data-set. We have several ideas on how to improve our approach to identifying the polarity in political tweets using information carried in Hashtags and Twitter users'name. The detection of irony and the study of re-tweet phenomena [38] can be two important elements to well classify tweets. In forthcoming works, we think to study in detail the impact of these phenomena in the micro-blogs classification.

6.1 Acknowledgment

This work is funded in part by the project ImagiWeb ANR-2012-CORD-002-01 (France).

References

1. Alonso Quecuty M-L. and Castillo M-D. *Detectando la ironía: La hipótesis aditiva como alternativa a las de referencia y la intención* In Cognitiva, Vol. 3, N^o 1, 1991, pp. 95-122
2. Amigó, E., Corujo, A., Gonzalo, J., Meij, E., de Rijke, M. *Overview of RepLab 2013: Evaluating Online Reputation Management Systems* CLEF 2013 Labs and Workshop Notebook Papers (2013).
3. Baccianella S., Esuli A., and Sebastiani F. *SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining* In Proceedings of the Seventh Conference on International Language Resources and Evaluation, pp. 2200-2204. European Language Resources Association.
4. Barbieri F. and Saggion H. *Modelling Irony in Twitter.* Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, EACL, April 26-30, 2014, Gothenburg, Sweden, pp. 56–64.
5. Barbieri F. and Saggion H. *Modelling Irony in Twitter: Feature Analysis and Evaluation* In Proceedings of Language Resources and Evaluation Conference (LREC) 2014 4258-4264
6. Bertrand, K. Z., Bialik, M., Virdee, K., Gros, A., and Bar-Yam, Y. *Sentiment in New York City: A High Resolution Spatial and Temporal View* arXiv preprint arXiv:1308.5010.
7. Bing Liu *Sentiment Analysis and Opinion Mining* Morgan & Claypool Publishers, May 2012.
8. Bollen J. Mao H. and P. A. *Modelling public mood and emotion: Twitter sentiment and socio-economic phenomena* In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011.
9. Carrillo-de-Albornoz, J., Amigó, E., Spina, D., and Gonzalo, J. *ORMA: A Semi-automatic Tool for Online Reputation Monitoring in Twitter* In Advances in Information Retrieval (pp. 742-745). Springer International Publishing.
10. Cha M., Haddadi H., Benevenuto F. and Gummadi K. P. *Measuring User Influence in Twitter: The Million Follower Fallacy* In Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM), Washington, DC, 2010.

11. Chung, J. E., and Mustafaraj, E. *Can collective sentiment expressed on Twitter predict political elections ?* In Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI 2011) pp. 1768-1769. Menlo Park, CA: AAAI Press.
12. Dodds, P. S., and Danforth, C. M. *Measuring the happiness of large-scale written expression: Songs, blogs, and presidents* Journal of Happiness Studies, 11(4), 441-456.
13. Dadidov D., Tsur O., and Rappoport A. *Semi-supervised recognition of sarcastic sentences in Twitter and Amazon* In Proceedings of the Fourteenth Conference on Computational Natural Language Learning. Association for Computational Linguistics, 2010. p. 107-116.
14. Fabbrizio GD, Tur G and Hakkani-Tür D 2004 *Bootstrapping spoken dialogue systems with data reuse* Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue.
15. Fellbaum C. (1998) *WordNet: An Electronic Database* Cambridge, MIT Press.
16. González-Ibáñez, R., Muresan, S. and Wacholder, N. *Identifying sarcasm in Twitter: a closer look* In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Short papers-Volume 2. Association for Computational Linguistics, 2011. p. 581-586.
17. Gruzd, A., Doiron, S., and Mai, P. *Is happiness contagious online? A case of Twitter and the 2010 Winter Olympics* In Proceedings of the 44th Hawaii International Conference on System Sciences. Washington, DC: IEEE Computer Society.
18. Kramer, A. D. I. *An unobtrusive behavioral model of "Gross National Happiness"* In Proceedings of CHI 2010 (pp. 287-290). New York: ACM Press.
19. Lampos V, Preotiuc-Pietro D, and Cohn T *A user-centric model of voting intention from Social Media* In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013).
20. Martínez, V., and González, V. M. *Sentiment Characterization of an Urban Environment via Twitter* In Ubiquitous Computing and Ambient Intelligence. Context-Awareness and Context-Driven Interaction (pp. 394-397). Springer International Publishing.
21. Maynard, D., Bontcheva, K., and Rout, D. *Challenges in developing opinion mining tools for social media* In Proceedings of NLP can u tag # user generated content, 2012.
22. Malouf R. and Mullen T. 2008. *Taking sides: User classification for informal online political discourse* In Internet Research, volume 18, pages 177-190.
23. O'Connor B., Balasubramanyan R., Routledge B. and Smith N. 2010. *From tweets to polls: Linking text sentiment to public opinion time series* In International AAAI Conference on Weblogs and Social Media.
24. Pang, B., and Lee, L. *Opinion mining and sentiment analysis. Foundations and Trends* In Information Retrieval 1(1-2), 1-135.
25. Pla F. and Hurtado L. *Political Tendency Identification in Twitter using Sentiment Analysis Techniques* In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers pp183-192, Dublin (Ireland)
26. Reyes A., Rosso P., and Veale T. *A Multidimensional Approach for Detecting Irony in Twitter.* In Language Resources and Evaluation, vol. 47, issue 1, pp. 239-268.
27. Saralegi X., San Vicente I. *Elhuyar at TASS 2013.* In Proceedings of "XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural". Workshop on Sentiment Analysis at SEPLN (TASS2013). Madrid. pp. 143-150.
28. Stone, P. J., Dexter C. D., Marshall S. S., and Ogilvie D. *The General Inquirer: A Computer Approach to Content Analysis.* Cambridge, MA: MIT Press.

29. Tello-Leal E., Tello-Leal D. A. and Sosa Reyna C.M. *Reflexiones sobre el uso de las tecnologías de información y comunicación en las campañas electorales en México: e-campañas.* Revista Virtual Universidad Católica del Norte, 36 (2012), 33-47.
30. Thelwall, M., Buckley, K., and Paltoglou, G. *Sentiment in Twitter events.* Journal of the American Society for Information Science and Technology, 62(2), 406-418.
31. Torres-Moreno, J.-M., El-Beze, M., Bellot, P. and Bechet, *Opinion detection as a topic classification problem* In in Textual Information Access. Chapter 9, pp 337, John Wiley & Son. 2013
32. Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. *Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment* ICWSM, 10, 178-185.
33. Velcin J., Kim Y.-M, Brun C., Dormagen J.-Y, SanJuan E., Khouas L., Peradotto A., Bonnevay S., Roux C., Boyadjian J., Molina A. and Neihouser M. *Investigating the Image of Entities in Social Media: Dataset Design and First Results* Proceedings of Language Resources and Evaluation Conference (LREC), 2014
34. Villena-Román J. and García-Morera, J. *TASS 2013-Workshop on Sentiment Analysis at SEPLN 2013: An overview.*
35. Wang H., Can D., Kazemzadeh A., Bar F. and Narayanan S. *A system for real-time twitter sentiment analysis of 2012 U.S. presidential election cycle* In Proceedings of the ACL 2012 System Demonstrations, ACL'12, pages 115-120. Association for Computational Linguistics
36. Wijaya, V., Erwin, A., Galinium, M., and Muliady, W. *Automatic mood classification of Indonesian tweets using linguistic approach* In 2013 International Conference on Information Technology and Electrical Engineering (ICITEE). pp. 41-46. IEEE.
37. Wilson T., Wiebe J., and Hoffmann P. *Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis.* In Proceedings of HLT-EMNLP-2005.
38. Morchid M., Dufour R., Bousquet P-M., Linarès G., Torres-Moreno, J-M. *Feature selection using Principal Component Analysis for massive retweet detection*, Pattern Recognition Letters ; DOI: 10.1016/j.patrec.2014.05.020