Overview of the 1st Classification of Spanish Election Tweets Task at IberEval 2017

Maite Giménez[1], Tomás Baviera[2], Germán Llorca[3], José Gámir[3], Dafne Calvo[4], Paolo Rosso[1], and Francisco Rangel[1,5]

[1] Pattern Recognition and Human Language Technology (PRHLT) Research Center, Universitat Politècnica de València
{mgimenez,prosso}@dsic.upv.es

[2] Mediaflows Research Group, Valencian International University
tomas.baviera@campusviu.es

[3] Mediaflows Research Group, Universitat de València
{german.llorca, jose.gamir}@uv.es

[4] Mediaflows Research Group, Universidad de Valladolid
dafne.calvo@uva.es

[5] Autoritas Consulting, S.A., Spain
francisco.rangel@autoritas.es

Abstract. This paper summarises the COSET shared task organised as part of the IberEval workshop. The aim of this task is to classify the topic discussed in a tweet into one of five topics related to the Spanish 2015 electoral cycle. A new dataset was curated for this task and hand-labelled by experts on the task. Moreover, the results of the 17 participants of the task and a review of their proposed systems are presented. In a second phase evaluation, we provided the participants with 15.8 millions tweets in order to test the scalability of their systems.

Keywords: Topic Classification, Twitter, Elections

1 Introduction

Nowadays, politics has upended by the usage of social media. A political campaign cannot be strategised using only the traditional media. During the election cycle, both politicians and voters engage in conversations about different topics. Politicians and their campaign staff share their policy approaches and bits of the candidates' personal lives. Characterising the influence processes in the public space is one of the most interesting topics in political communication research. Political parties, media and citizens send messages through a complicated media network, where knowing who has the power of agenda setting becomes critical. In this sense, the social media logic has boosted a more active user participation in delivering political messages, accessing more sources, and mobilising for political action. The analysis of this complex media network requires innovative research tools capable of evaluating the different elements in the political information flow [5].

To create a shared framework, we have proposed a shared task: the Classification of Spanish Election Tweets (COSET) task, which tackles the problem of topic classification of political tweets in five categories.

The political background of the COSET project was one of the most uncertain electoral contests in the Spain's recent political history: the December 20, 2015 General Elections. The European Elections of the previous year had consolidated two new parties in the national political landscape. Both sought to challenge the bipartisanship entrenched in Spanish democracy. For the 2015 General Elections, the campaign uncertainty, as well as the increased number of candidates with possibilities of success, made the citizenry more interested in the campaign than ever in recent history. The traditional media, particularly TV, and social media widely covered politics during the weeks prior to Election Day [23].

The remainder of this paper is organised as follows. Section 2 illustrates the state of the art on the topic. Following, Section 3 describes the corpus and the process for collecting the tweets from the political conversations on Twitter related to the 2015 Spanish General Elections, as well as the evaluation framework proposed for evaluating the participants' models. Section 4 summarises the proposed approaches submitted by the participants, and the results achieved by the models evaluated are discussed. Finally, in Section 5 the conclusions are presented.

2 Related Work

The following sections describe the work related to topic classification as well as the work of Natural Language Processing (NLP) in political campaigns.

2.1 Topic Classification Using Natural Language Processing

Topic classification is one of the classical problems of NLP. In the literature, we find that this task has been tackled following a wide variety of approaches.[6] The task at hand has been studied in depth because it can be used as a first step for extracting relevant information from a text [18]. The work of Hillard et al. [15], depicts an example of how automatic classification systems can assist human annotators in labelling the topic discussed in a document. In a structured text, the state of the art has achieved satisfactory results in most domains. However, this task can be challenging when dealing with the short texts with many grammatical mistakes found on social media [21, 36, 6]. Furthermore, recently social media has been used extensively during the elections, which has aroused the interest of researchers working both on computational linguistics and social science studies [20, 12, 42].

Content classification of tweets in political research has been addressed mainly on lexicon-based methods. A previous issue selected from the campaign provides

[6] For more information, please review the survey that can be found in the following reference [1, chap. 6]

the list of topics per which tweets will be classified [8]. This method has also been used for identifying political influencers [10]. Other classifications use methods based on network graphs for uncovering word patterns [37]. Moreover, these works have explored the impact of different machine learning algorithms in order to predict the output of the elections (e.g. Support Vector Machines (SVMs) [7], Linear Discriminant Analysis (LDA) [31], etc.) Likewise, some works linked the output of the election with the sentiments expressed on Twitter [39, 41, 38].

The utility of these methodologies relies on the set of words that distinguish among the topics, such as *economy* or *national security*. Nevertheless, these methods miss critical issues within the political conversation as they usually focus on sectorial policies. To address the broader spectrum of political topics discussed on Twitter, researchers need to develop more refined machine-learning-based methods able to detect more abstract topics.

2.2 Topic Labelling in Political campaigns

To label the data set that we have collected, we followed the topic classification proposed by Mazzoleni [26], as this is the baseline for the content analysis carried out by the entire Mediaflows research project. Patterson [28] distinguishes among four kinds of basic issues present in the media during the campaign. Mazzoleni [26] assumes this taxonomy in his studies on mediatised politics.

According to Patterson [28], the media's messages during the campaign fall into four categories based on their political content[7]: (i) political issues, dealing with the most abstract aspects of electoral confrontation; (ii) policy issues, dealing with sectorial policies; (iii) personal issues, regarding the candidates' lives and pastimes and; (iv) campaign issues, dealing with the evolution of the campaign. Although we had set some filtering criteria in the process of extraction, we may have collected some tweets unrelated to the Spanish Elections or the political campaign. Thus, we decided to introduce a fifth category (v) other issues for this kind of content.

3 Evaluation Framework

This section defines the task at hand, outlines the construction of the corpus highlighting the annotation process details, and describes the performance metric used to evaluate the participants' models.

3.1 Corpus: Tweet Collection and Annotation

In order to carry out this task, we gathered a collection of tweets from November 2, 2015, to December 21, 2015. Of these 50 days, 32 correspond with the pre-campaign, 15 with the electoral campaign, one with reflection day, one with Election Day, and one more with the following day. This last day is useful because

[7] `http://mediaflows.es/coset/`

the conversations after knowing the results on Election Day ended at midnight. The tweets were obtained through the Twitter API. The data mining and the pre-processing of tweets were conducted using Python.

We established three criteria for filtering tweets: a pair of general terms related to the elections (*#20D; 20-D*); the names of the four major political parties along with their Twitter handles (PP; PPopular; PSOE; @PSOE; ahorapodemos; Ciudadanos; CiudadanosCs; Cs); and the names of the four prime minister candidates along with their Twitter handles (Rajoy; @marianorajoy; Pedro Sanchez; Pedro Snchez, @sanchezcastejon; Pablo Iglesias; @Pablo_Iglesias_; Rivera; Albert_Rivera). It was impossible to include the name of the political party Podemos as a filter element. This word works poorly in constructing a corpus through a selective extraction process because, given that it means *we can*, it can be used in many contexts other than political conversations. We also filtered out messages written in languages other than Spanish.

3.2 Task definition

As we establish in the Introduction, currently, political campaigns monitor political conversations on Twitter, particularly when an electoral cycle is approaching. This task is usually carried out in a semi-automatic fashion. The focus of the proposed task COSET is on improving this process. Therefore, participants were asked to classify tweets written in Spanish based on the political topic discussed. As mentioned in Section 2.2, we considered five categories:

1. Political Issues (PI): Tweets related to the most abstract elements of electoral confrontation.
2. Policy Issues (PoI): Tweets about sectorial policies.
3. Campaign Issues (CI): Tweets related to the evolution of the campaign.
4. Personal Issues(PeI): The candidates' personal lives and pastimes.
5. Other Issues (O): The tweets that did not fit in any of the previous categories.

Summarising, the objective of the task is when supplied with a tweet, the system proposed should be able to predict the tweet's topic automatically.

Participants were provided with password-protected labelled data sets for training and developing their systems. Later, their systems were evaluated against a test data set. Table 1 presents the distribution of tweets for each topic and data set, and Figure 1 shows the distribution of the topics over the whole dataset (including the training, testing, and developing partitions)

3.3 Performance measures

Given that the corpora were heavily unbalanced, as we have illustrated in the previous section, we proposed ranking the participants' models using the macro F_1-score. The F-score can be interpreted as a weighted average of the precision

Table 1. Distribution of the number of tweets for each topic and data set.

	Training	Development	Testing
PI	530 (23.64 %)	57 (22.8 %)	151 (24.2 %)
PoI	786 (35.06 %)	88 (35.2 %)	228 (36.54 %)
CI	511 (22.79 %)	71 (28 %)	136 (21.79%)
PeI	152 (6.78 %)	9 (4 %)	38 (6.09%)
O	263 (11.73 %)	25 (10 %)	71 (11.38%)
Total	2242	250	624

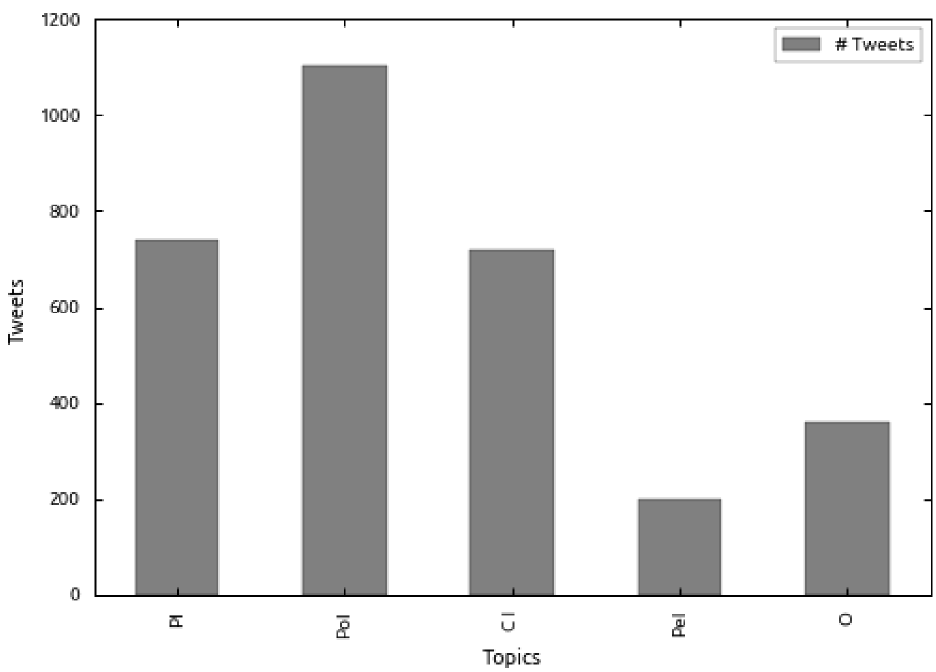

Fig. 1. Distribution of the number of tweets for each topic in the dataset labelled.

and the recall, whereas the F_1-score is the harmonic mean of the precision and recall metrics as seen in Formula 1.

$$\begin{aligned} F_1 &= 2 \cdot \frac{precision \cdot recall}{precision + recall} \\ precision &= \frac{1}{|L|} \sum_{l \in L} Pr(y_l, \hat{y}_l) \\ recall &= \frac{1}{|L|} \sum_{l \in L} R(y_l, \hat{y}_l) \end{aligned} \tag{1}$$

where $|L|$ is the number of samples, $\hat{y}_l$ is the true label for the sample l, and y_l is the predicted label for the sample l [34].

Facing multi-class tasks, we also need to take into account the weighted average of the F_1-score of each class. Since we wanted to penalise those systems that have bias towards the most populated classes, we have used the macro average, which calculates the unweighted mean for each label as described in Formula 2.

$$F_{1-macro} = \frac{1}{|L|} \sum_{l \in L} F_1(y_l, \hat{y}_l) \tag{2}$$

4 Overview of the Submitted Approaches

Hereafter, we present a summary of the proposed models as well as the results that each model achieved. We should note that, each participant was allowed to submit up to five different proposals in order to allow them to test different approximations. In total, 17 teams participated in the task, and a total of 39 models were submitted.

Pre-process Most of the participants did not pre-process the tweets from the data sets and worked with the raw data. However, the techniques used for those who did pre-process the data sets were: tokenisation (carried out by teams LuSer[4], Carl Os Duty [9], UC3M [11], and ivsanro1 [32]) conversion to lowercase (teams LuSer [4], UC3M [11], and Electa[16]), and removal of several tokens such as user handles (teams LuSer [4], ELiRF-UPV [14], and slovak [24]), numbers (teams ELiRF-UPV [14], and slovak [24]), punctuation marks (teams Electa[16], slovak [24], and ivsanro1 [32]), URLs (teams Electa[16], slovak [24], and ivsanro1 [32]), stopwords (teams Electa[16], UC3M [11], and slovak [24]), flooding characters (team slovak [24], and UC3M [11]), and emoticons (team Electa[16]).

Features The features used to train the participants' classifiers were diverse. Participants' models used some classical features in NLP such as word n-grams (teams LuSer [4], LTRC_IIITH [19], ConradCR [3], Electa [16], Team 17 [40], Carl Os Duty [9], Citripio [25], LichtenwalterOlsan [22], slovak [24], Puigcerver [30], and ivsanro1 [32]), character n-grams (team LTRC_IIITH [19]), Tf-Idf (teams CD_team [33], Carl Os Duty [9], LichtenwalterOlsan [22], and Puigcerver [30]); but some of them used more recent techniques such as word embeddings (teams LTRC_IIITH [19], ELiRF [14], atoppe [2], UC3M [11], and MíVal [27]), sentence embeddings (Team 17 [40]), and a multi-dimensional vector approach (team UT text miners [13]). Moreover, the work of LTRC_IIITH [19] used an extensive set of handcrafted features that included top tokens, hashtags, hashtag decomposition, mentions, and URLs among others.

Classification approaches The most used model for addressing the task was a model based on Neural Networks (NNs) (teams LTRC_IIITH [19], ELiRF [14], Team 17 [40], and UT text miners [13]); LuSer [4] added normalisation techniques such as Gaussian Noise to the NNs architecture, and Carl Os Duty [9] included batch normalisation with dropout in their NN model. In addition, other approaches were also considered such as Support Vector Machines (teams LTRC_IIITH [19], MíVal [27], and Citripio [25]), Random Forests (teams LTRC_IIITH [19], ConradCR [3], and Electa [16]), Naïve Bayes (teams slovak [24] and ivsanro1 [32]), Logistic Regression (team Puigcerver [30]); CD_team [33] proposed a combination of classifiers that included a Logistic Regression, an SVM, Naive Bayes, and a K-Nearest Neighbours classifier. Deep learning models were also considered in the work of team atoppe [2]; they experimented with Convolutional Neural Networks, Long Short Term Memory Networks (LSTMs), Bidirectional Long Short-Term Memory Networks, etc. Also, team UC3M [11] addressed this task using LSTMs, and Gated Recurrent Units. Furthermore, team 17 [40] trained five different language models for each topic and then classified each tweet minimising the perplexity of language models.

4.1 Evaluation and Discussion of the Submitted Approaches

First, we have developed three baselines to meet different difficulty levels. The first baseline is the simplest one, and it will always predict the most common class Policy Issues (PI). The second is a traditional machine learning approach that uses a Bag of Words (BOW) and an SVM with a linear kernel. Finally, the last baseline proposed applies a slightly better representation of words following a term frequency–inverse document frequency (Tf-idf) [17] and Random Forests (RF) for classifying the training samples. None of these baselines has its hyperparameters adjusted to fit the task, and they were developed using the Scikit-learn package [29]. The results of all the participants' models are presented in Table 2.

Overall, this is a complicated task since several topics are similar and, therefore, share parts of the vocabulary. Only the first ten systems are able to achieve an F_1 macro over 0.6. The best result was obtained by ELiRF-UPV [14], who used NNs and word embeddings to train their systems, but also included a technique for handling the imbalance present in the data. Also, LuSer [4] applied NNs, but in this case, they used 3-grams as features and included Gaussian Noise, which is reported to help to minimize the effect of overfitting in NNs. It is worth noting that some systems were unable to improve the results achieved by some of the baseline systems.

We have studied the confusion matrix of the three best-performing systems, the first and fourth runs from the ELiRF-UPV [14] team and the run from the

Table 2. Distribution of tweet for each topic and data set. Where Tf-idf: term frequencyinverse document frequency; RF: random forest, SVM: Support Vector Machine; and BOW: bag of words.

Team	run	F_1 macro
ELiRF-UPV	**run 1**	**0.6482**
ELiRF-UPV	run 4	0.6400
LuSer	run 1	0.6337
ELiRF-UPV	run 3	0.6330
ELiRF-UPV	run 2	0.6233
Puigcerver	run 1	0.6176
atoppe	run 3	0.6157
atoppe	run 2	0.6065
LTRC_IIITH	run 2	0.6054
LTRC_IIITH	run 4	0.6049
Puigcerver	run 2	0.5997
LTRC_IIITH	run 3	0.5960
LTRC_IIITH	run 1	0.5959
atoppe	run 5	0.5952
Carl Os Duty	run 1	0.5902
CD_team	run 1	0.5859
MVal	run 2	0.5852
Carl Os Duty	run 2	0.5822
Electa	run 1	0.5784
atoppe	run 1	0.5745
MíVal	run 1	0.5733
Citripio	run 1	0.5676
ConradCR	run 1	0.5639
LichtenwalterOlsan	run 1	0.5590
UT text miners	run 3	0.5541
atoppe	run 4	0.5476
Puigcerver	run 3	0.5275
ivsanro1	run 1	0.5234
LTRC_IIITH	run 5	0.4435
Baseline: Tf-idf & RF	-	0.4236
slovak	run 1	0.4233
UT text miners	run 1	0.3631
UT text miners	run 2	0.3341
UC3M	run 3	0.2755
UC3M	run 4	0.2755
Baseline: BOW & SVM	-	0.2644
UC3M	run 5	0.2615
UC3M	run 1	0.2571
UC3M	run 2	0.2558
Team 17	run 2	0.2446
Team 17	run 1	0.241
Baseline: Most frequent	-	0.107

team LuSer [4], which corresponds with Figures 2, 3, and 4 respectively. It can be observed that the predictions made for the topics PI, PoI, and CI present certain confusion between them. Remarkably, PoI is the easiest topic to classify. In contrast, the topic PeI is the most challenging.

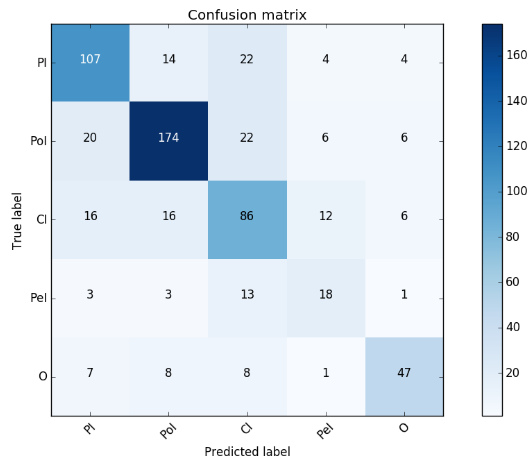

Fig. 2. Confusion matrix for the run 1 from ELiRF team

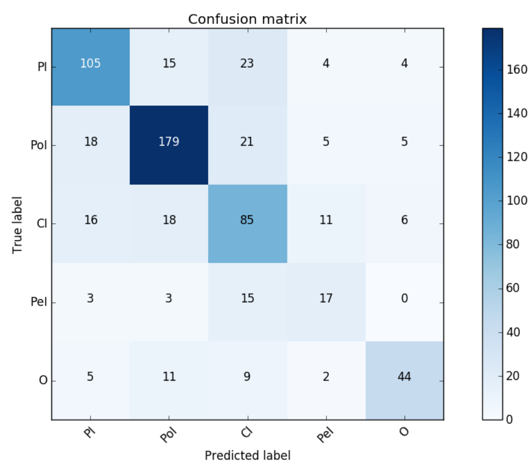

Fig. 3. Confusion matrix for the run 4 from ELiRF team

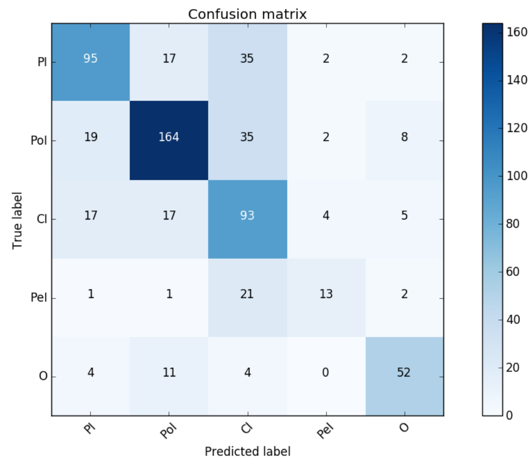

Fig. 4. Confusion matrix for the run 1 from LuSer team

4.2 Second Phase Evaluation

We have offered the participants the opportunity to test the scalability of their approaches with a bigger dataset of 15.8 millions tweets. Being practically impossible to manually label such a large corpus, we have built a silver standard with pooling techniques [35]. Four were the teams who submitted their runs. The best performing team [14] submitted two runs and the other teams [2, 40, 19] submitted one run each. We have prepared a pool formed by these five runs and labelled the corpus with the agreement of at least four runs (80% of agreement). The corpus size before and after labelling, besides the distribution of labels, is shown in Table 3. As can be seen, the labelled corpus with the agreement of three runs comprises 65.91% of the original corpus.

Table 3. Corpus size before and after labelling, and distribution of labels (using the pooling technique).

Corpus	**Size**	**Percentage**
Complete	15,806,058	-
Labelled	10,417,058	65.91%
Label	**Size**	**Percentage**
PI	2,153,236	20.67%
PoI	3,732,610	35.83%
CI	3,127,160	30.02%
PeI	581,089	5.58%
O	822,963	7.90%

In Table 4, results for the second phase are shown. As can be seen, the best performing team also obtains the highest F_1 value. On the contrary, Team 17 has increased its performance due to the use of fastText in this second phase evaluation.

Table 4. Results in the second phase evaluation in terms of F_1 macro.

Team	F_1 macro
ELiRF-UPV.1	0.9586
ELiRF-UPV.2	0.9523
Team 17	0.9482
atoppe	0.8960
LTRC_IIITH	0.8509

5 Conclusions

This paper summarises the first edition of the task COSET on topic classification during the 2015 electoral cycle. COSET was one of the tasks from the IberEval workshop, which was part of the annual Conference held by the Spanish Society for Natural Language Processing (SEPLN in Spanish). Given a set of tweets, participants were asked to classify the topic discussed in them from a list of five topics that included: political issues, policy issues, campaign issues, personal issues, and other issues. Seventeen participants performed the task, and the best result was achieved by *ELiRF-UPV* [14] who scored 0.6482 in the F_1 macro. They applied NNs, word embeddings, and handled the imbalance present in the data. The results achieved by the participants confirm that topic classification from tweets is a difficult task, particularly when the topics are similar. Hence, a shared task for evaluating different systems, like the ones proposed in this task, can help improve the results of automatic classification or at least assist human labelling. This has been the aim of the second phase evaluation.

Acknowledgments

This work was conducted under the auspices of the CSO2016-77331-C2-1-R research project "Strategies, agendas and discourses in the electoral cybercampaigns: media and citizens" (Mediaflows), funded by the Spainish Ministry of Economy, Industry and Competitiveness (MINECO in Spanish), and under the the auspices of the TIN2015-71147-C2-1-P research project "SOcial Media language understanding-EMBEDing contexts" (SomEMBED), funded by MINECO. The work of the first author is financed by Grant PAID-01-2461 2015, from the Universitat Politècnica de València.

References

1. Aggarwal, C.C., Zhai, C.: Mining text data. Springer Science & Business Media (2012)
2. Ambrosini, L., Nicolò, G.: Comparative study of neural models for the COSET shared task at IberEval 2017. In: Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017). CEUR Workshop Proceedings. CEUR-WS.org, Murcia (Spain) (September 19 2017)
3. Bernath, C.: Submission to the 1st classification of spanish election tweets task at ibereval 2017. `http://mediaflows.es/coset/`. Team: ConradCR. Spain
4. Cebrián Chuliá, L., Ferrer Sánchez, S.: Classification Of Spanish Election Tweets (COSET) with neural networks. In: Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017). CEUR Workshop Proceedings. CEUR-WS.org, Murcia (Spain) (September 19 2017)
5. Chadwick, A.: The hybrid media system: Politics and power. Oxford University Press (2013)
6. Chen, Y., Li, Z., Nie, L., Hu, X., Wang, X., Chua, T.s., Zhang, X.: A semi-supervised bayesian network model for microblog topic classification. In: Coling. pp. 561–576 (2012)
7. Conover, M.D., Gonçalves, B., Ratkiewicz, J., Flammini, A., Menczer, F.: Predicting the political alignment of twitter users. In: Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. pp. 192–199. IEEE (2011)
8. Conway, B.A., Kenski, K., Wang, D.: The rise of twitter in the political campaign: Searching for intermedia agenda-setting effects in the presidential primary. Journal of Computer-Mediated Communication 20(4), 363–380 (2015)
9. Diez Alba, C., Vieco Pérez, J.: IberEval 2017, COSET task: a basic approach. In: Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017). CEUR Workshop Proceedings. CEUR-WS.org, Murcia (Spain) (September 19 2017)
10. Dubois, E., Gaffney, D.: The multiple facets of influence: identifying political influentials and opinion leaders on twitter. American Behavioral Scientist 58(10), 1260–1277 (2014)
11. Fernandez Hernandez, A., Segura Bedmar, I.: Submission to the 1st classification of spanish election tweets task at ibereval 2017. `http://mediaflows.es/coset/`. Team: UC3M. Spain
12. Gayo Avello, D., Metaxas, P.T., Mustafaraj, E.: Limits of electoral predictions using twitter. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. Association for the Advancement of Artificial Intelligence (2011)
13. Gharavi, E., Bijari, K.: Short text classification using deep representation: A case study of Spanish tweets in COSET Shared Task. In: Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017). CEUR Workshop Proceedings. CEUR-WS.org, Murcia (Spain) (September 19 2017)
14. González, J.A., Pla, F., Hurtado, L.F.: ELiRF-UPV at IberEval 2017: Classification Of Spanish Election Tweets (COSET). In: Proceedings of the Second Workshop

on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017). CEUR Workshop Proceedings. CEUR-WS.org, Murcia (Spain) (September 19 2017)

15. Hillard, D., Purpura, S., Wilkerson, J.: Computer-assisted topic classification for mixed-methods social science research. Journal of Information Technology & Politics 4(4), 31–46 (2008)
16. Juárez, G., Peralta, A.: Submission to the 1st classification of spanish election tweets task at ibereval 2017. `http://mediaflows.es/coset/`. Team: Electa. Spain
17. Jurafsky, D., Martin, J.H.: Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition
18. Kao, A., Poteet, S.R.: Natural language processing and text mining. Springer Science & Business Media (2007)
19. Khandelwal, A., Swami, S., Akhtar, S.S., Shrivastava, M.: Classification Of Spanish Election Tweets (COSET) 2017: Classifying Tweets using Character and Word Level Features. In: Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017). CEUR Workshop Proceedings. CEUR-WS.org, Murcia (Spain) (September 19 2017)
20. Larsson, A.O., Moe, H.: Studying political microblogging: Twitter users in the 2010 swedish election campaign. New Media & Society 14(5), 729–747 (2012)
21. Lee, K., Palsetia, D., Narayanan, R., Patwary, M.M.A., Agrawal, A., Choudhary, A.: Twitter trending topic classification. In: Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on. pp. 251–258. IEEE (2011)
22. Lichtenwalter, D., Olšan, T.: Submission to the 1st classification of spanish election tweets task at ibereval 2017. `http://mediaflows.es/coset/`. Team: LichtenwalterOlsan. Czech Republic and Germany
23. López García, G., Valera Ordaz, L.: Pantallas electorales. El discurso de partidos, medios y ciudadanos en la campaña de 2015. Editorial UOC (2017)
24. Mahiques Sifres, X., Lyeuta Tykhovod, V.: Submission to the 1st classification of spanish election tweets task at ibereval 2017. `http://mediaflows.es/coset/`. Team: slovak. Spain
25. Maluenda Maez, F., Garca Ferrando, G.A.: Submission to the 1st classification of spanish election tweets task at ibereval 2017. `http://mediaflows.es/coset/`. Team: Citripio. Spain
26. Mazzoleni, G.: La comunicación política. Alianza Editorial (2014)
27. Míguez, O., Valdiviezo, M.: Submission to the 1st classification of spanish election tweets task at ibereval 2017. `http://mediaflows.es/coset/`. Team: MíVal. Spain
28. Patterson, T.: The Mass Media Election: How Americans Choose Their President., vol. 75. New York: Praeger Special Studies. (1980)
29. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830 (2011)
30. Puigcerver, J.: Submission to the 1st classification of spanish election tweets task at ibereval 2017. `http://mediaflows.es/coset/`. Team: Puigcerver. Spain
31. Quercia, D., Askham, H., Crowcroft, J.: Tweetlda: supervised topic classification and link prediction in twitter. In: Proceedings of the 4th Annual ACM Web Science Conference. pp. 247–250. ACM (2012)
32. Sánchez, I.: Submission to the 1st classification of spanish election tweets task at ibereval 2017. `http://mediaflows.es/coset/`. Team: ivsanro1. Spain

33. De la Peña Sarracén, G.L.: Ensembles of methods for Tweet Topic Classification. In: Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017). CEUR Workshop Proceedings. CEUR-WS.org, Murcia (Spain) (September 19 2017)
34. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Information Processing & Management 45(4), 427–437 (2009)
35. Spark-Jones, K.: Report on the need for and provision of an'ideal'information retrieval test collection. Computer Laboratory (1975)
36. Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., Demirbas, M.: Short text classification in twitter to improve information filtering. In: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. pp. 841–842. ACM (2010)
37. Sudhahar, S., Veltri, G.A., Cristianini, N.: Automated analysis of the us presidential elections using big data and network analysis. Big Data & Society 2(1) (2015)
38. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Computational linguistics 37(2), 267–307 (2011)
39. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with twitter: What 140 characters reveal about political sentiment. ICWSM 10(1), 178–185 (2010)
40. Villar Lafuente, C., Garcés Díaz-Munío, G.: Several approaches for tweet topic classification in COSET - IberEval 2017. In: Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017). CEUR Workshop Proceedings. CEUR-WS.org, Murcia (Spain) (September 19 2017)
41. Wang, H., Can, D., Kazemzadeh, A., Bar, F., Narayanan, S.: A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In: Proceedings of the ACL 2012 System Demonstrations. pp. 115–120. Association for Computational Linguistics (2012)
42. Zirn, C., Glavaš, G., Nanni, F., Eichorts, J., Stuckenschmidt, H.: Classifying topics and detecting topic shifts in political manifestos. PolText (2016)