

# Overview of PoliticES at IberLEF 2023: Political Ideology Detection in Spanish Texts

## *Resumen de la tarea PoliticES en IberLEF 2023: Detección de Ideología Política en Español*

José Antonio García-Díaz,<sup>1</sup> Salud María Jiménez-Zafra,<sup>2</sup>  
 María-Teresa Martín Valdivia,<sup>2</sup> Francisco García-Sánchez,<sup>1</sup>  
 L. Alfonso Ureña-López,<sup>2</sup> Rafael Valencia-García<sup>1</sup>

<sup>1</sup>Facultad de Informática, Universidad de Murcia

<sup>2</sup>Computer Science Department, SINAI, CEATIC, Universidad de Jaén

{joseantonio.garcia8, frgarcia, valencia}@um.es

{sjzafra, maite, laurena}@ujaen.es

**Abstract:** This paper describes PoliticES 2023, a shared task organized within the workshop IberLEF 2023 in the framework of the 39th edition of the International Conference of the Spanish Society for Natural Language Processing. This second edition of the task shares the goal of the first edition of PoliticES, which is to extract political ideology and other psychographic and demographic characteristics of users in social networks. What is new this year is that the traits are extracted from text clusters of users who share the same traits, and that celebrities have been included as a type of profession. This edition attracted 43 teams, of which 11 submitted results and 8 sent papers describing their systems. Most of the participants proposed Transformers-based approaches, but others also used traditional machine learning algorithms.

**Keywords:** Cluster profiling, political ideology, author analysis, demographic and psychographic traits.

**Resumen:** Este artículo describe PoliticES 2023, una tarea organizada dentro del taller IberLEF 2023 en el marco de la 39 edición del Congreso Internacional de la Sociedad Española para el Procesamiento del Lenguaje Natural. Esta segunda edición de la tarea comparte el objetivo de la primera edición de PoliticES, extraer la ideología política y otros rasgos psicográficos y demográficos de usuarios en redes sociales. Las novedades son que este año los rasgos se extraen de clústers de textos de usuarios que comparten los mismos rasgos y que se ha incluido celebridades como tipo de profesión. Esta edición ha atraído a 43 equipos, de los cuales 11 enviaron resultados y 8 presentaron artículos describiendo sus sistemas. La mayoría de los participantes propusieron enfoques basados en Transformers, pero también otros utilizaron algoritmos tradicionales de aprendizaje automático.

**Palabras clave:** Perfilado de clústers, ideología política, análisis de autores, rasgos demográficos y psicográficos.

### 1 Introduction

Political ideology encompasses a range of psychological characteristics that contribute to our understanding of individual and collective behaviour. These traits include moral and ethical values as well as inherent attitudes, evaluations, biases and prejudices (Verhulst, Eaves, and Hatemi, 2012). A study conducted by Fatke (2017) showed a correlation between personality traits and political ideology. The author collected data from 21 countries and found correlations be-

tween the big five personality traits and political ideology. Specifically, conscientiousness showed a strong association with right-wing ideology, while openness to experience and agreeableness were significantly more associated with left-wing ideology. Furthermore, our political ideology has a significant impact on our daily lives. For example, Baumgaertner, Carlisle, and Justwan (2018) discovered a relationship between political ideology and citizens' attitudes towards vaccination campaigns against infectious diseases.

This is the second edition of the shared task PoliticES (García-Díaz et al., 2022). This new edition of PoliticES is held in the framework of IberLEF 2023 (Jiménez-Zafra, Rangel, and Montes-y Gómez, 2023), as part of the SEPLN 2023 conference, and its objective is to extract information about political ideology from clusters of texts. For this purpose, a cluster profiling task is proposed. It focuses on the identification of gender, profession and political spectrum from a binary and multi-class perspective. The novelty this year is that instead of profiling users, participants will work with clusters of texts written by different users but with the same characteristics, in order to avoid legal and ethical problems.

The previous edition of PoliticES (García-Díaz et al., 2022) attracted 20 teams which submitted results and, finally, 14 participants sent working notes for the description of their systems. In this first edition, participants had to identify 2 demographic traits, gender and profession, and political ideology as a psychographic trait in two axes, binary and multiclass. The novelties this year are that we have included users who are celebrities and that the tweets are organized in clusters of users who share the same traits.

The paper is structured as follows. Section 2 provides a description of the PoliticES 2023 shared task. Section 3 presents the dataset that was used in the competition. The participant approaches are summarized in Section 4, followed by the presentation of results and a discussion in Section 5. Lastly, Section 6 concludes the paper by offering insights and suggestions for future work.

## 2 Task description

The aim of the PoliticES 2023 shared task is to extract two demographic traits (gender and profession) and one psychographic trait (political ideology) from a cluster of tweets from users who share these characteristics. Political ideology is approached as both a binary problem (pib) and a multiclass problem (pim).

The categories for each trait are:

- gender: male, female.
- profession: political, journalist, celebrity.
- pib: left, right.

- pim: left, moderate left, right, moderate right.

The challenges involved in this shared task are:

1. Extraction of political ideology from a collection of texts. This is the second edition of the first Spanish collaborative task focused on political ideology. In addition, extracting data from clusters of texts rather than writings by the same author will make this task more difficult, as certain biases of the datasets will disappear.
2. Multi-class classification. The task of automatic document classification should be approached from a binary and multiclass perspective with four different classes.

The competition was organized through CodaLab and is available at the following link: <https://codalab.lisn.upsaclay.fr/competitions/10173>. It was divided into 3 phases: Practice, Evaluation and Post-Evaluation.

In the practice phase, participants were given a subset of the training data to familiarize themselves with the data format. In addition, they were provided a notebook containing a baseline system based on BoW (Bag-of-Words) and Logistic Regression to give a starting point for system development. They were then given the full training set to develop their approaches. Participants were allowed to make a maximum of 100 submissions to CodaLab. In the evaluation phase, the test partition was made available for the participants to label using the developed systems. This partition was used to evaluate the teams. They were allowed to make a maximum of 10 submissions through CodaLab, from which each team had to select the best one for the ranking. The ranking was determined using the arithmetic mean of the macro F1 scores for gender, profession, binary political ideology, and multi-class political ideology.

## 3 Dataset

The dataset released for this shared task follows the guidelines defined in the PoliCorpus 2020 dataset (García-Díaz, Colomo-Palacios, and Valencia-García, 2022) and the corpus

used for the PoliticES 2022 shared task (García-Díaz et al., 2022).

The tweets were compiled between 2020 and 2022 from the Twitter accounts of politicians, political journalists and celebrities in Spain using the UMUCorpusClassifier (García-Díaz et al., 2020). These users were selected because their political affiliation can be guessed from the party the politicians belong to, the editorial line of the newspapers the journalists write for, or the type of political parties the celebrities support.

The politicians were selected from: (1) members of the Spanish government, (2) members of the Spanish Congress and Senate, (3) mayors of some important cities in Spain, (4) presidents of autonomous communities, (5) former politicians, and (6) employees of political parties. The journalists were selected from different Spanish news media, such as ABC, El País, ElDiario, El Mundo or La Razón, among others. They are relevant figures in political journalism, and some of them publish opinion articles and participate in television debates, so their political orientation can be assumed with some guarantees. For the celebrities, we selected those that publicly expressed their political ideology or support for different political parties on social networks.

We removed tweets that shared content from news sites without using retweets. We also discarded those that contained mentions of news sites or some linguistic cues, such as the pipe symbol commonly used by news sites to categorize their news. We also anonymized Twitter mentions of politicians by replacing them with the token `@user`. In addition, mentions of other Twitter accounts were also coded as `@user`. Other entities, such as references to political parties, are also replaced by the token `@political_party`. As a result, the textual characteristics cannot be trivially guessed by reading the user’s name and searching for information about them on the web.

The next step was to organize the users into text clusters by mixing some of these extracted tweets to avoid ethical and privacy issues with author profiling on Twitter. Each cluster is composed of 80 tweets written by different users who share all the characteristics under evaluation. We labeled each cluster with self-assigned gender (male, female), profession (celebrity, politician, jour-

nalist), and political spectrum on two axes: binary (left, right) and multiclass (left, moderate\_left, moderate\_right, right).

The final dataset consists of 2797 clusters of 80 tweets each. For the shared task, training and test sets were released (80%-20%). We released the dataset in two parts: training and testing. However, in the first phase of the competition, we released an early bird dataset consisting of a subset of 360 clusters for training and 90 for testing, each with 40 tweets. It is worth noting that the accounts from training and testing are completely independent to avoid automatic classifiers learning to identify the authors rather than the features. The number of users per set and feature is shown in Table 1.

#### 4 Participant approaches

The PoliticES 2023 shared task attracted 43 teams that registered in CodaLab, of which 11 submitted results and 8 presented working notes describing their systems. Below is a brief summary of the participants’ proposals:

- **(1st) ELiRF-VRAIN** (Ahuir et al., 2023). The ELiRF-VRAIN team approached the competition using two Spanish Transformers: MarIA (Gutiérrez Fandiño et al., 2022) and BETO (Cañete et al., 2020). They divided the dataset into batches, as all tweets from a cluster does not fit into a Transformer architecture due to max length restrictions, and performed the classification using a voting scheme based on the mode. The Transformers were fine-tuned using hyperparameter tuning. They also propose IDC (Ideology Discrepancy Correction) to align discrepancies between the binary and multiclass classifiers for political ideology. They achieve the 1st position in the ranking and the 1st position in the two binary classification tasks: gender and binary political ideology. They achieve the 3rd position for the identification of the profession and the 2nd position for the identification of political ideology in a multiclass axis.
- **(2nd) HiTZ-Ixa** (Fernandez de Landam and Agerri, 2023). This team proposes a multi-level representation

Trait	Labels	Train	Test	Total
Gender	Female	757	166	<b>923</b>
	Male	1493	381	<b>1874</b>
Profession	Celebrity	113	56	<b>169</b>
	Journalist	1385	305	<b>1690</b>
	Politician	752	186	<b>938</b>
Binary ideology	Left	1255	327	<b>1582</b>
	Right	995	220	<b>1215</b>
Multiclass ideology	Left	430	117	<b>547</b>
	Moderate left	825	210	<b>1035</b>
	Moderate right	728	153	<b>881</b>
	Right	267	67	<b>334</b>

Table 1: Corpus statistics per trait and labels.

to combine both user-level and tweet-level representations derived from textual data. For this, they evaluate several Transformers, including multilingual BERT (Devlin et al., 2019), DistilmBERT (Sanh et al., 2019), XLM-RoBERTa (Conneau et al., 2020), XLM-T (Barbieri, Espinosa Anke, and Camacho-Collados, 2022), mDeBERTa (He, Gao, and Chen, 2021), BETO (Cañete et al., 2020), PolitiBETO (Villa-Cueva et al., 2022), and MarIA (Gutiérrez Fandiño et al., 2022). Then, they extract the sentence embeddings of each model, and average all representations of each cluster to create an user representation. Next, they concatenate each tweet with its user representation. These features are the input of four Logistic Regression classifiers, one for each of the features. They achieve the 2nd position in the ranking, but reach the 1st position for the identification of the users’ profession.

- **(3rd) ESCOM-IPN** (Acosta-Pacheco et al., 2023). They achieve the 3rd position using traditional machine learning classifiers like Logistic Regression or SVMs with statistical features including term frequency and TF-IDF. First, they concatenate all tweets in each cluster, next they apply tokenization, lemmatization, and clean the texts by removing stop words and punctuation. These methods proved to be very competitive for political classification, reaching the 2nd position in the binary task and the

1st position in the multiclass labeling. This result suggests that the identification of political ideology can be done more easily by identifying certain keywords in the documents.

- **(4th) INGEOTEC**. This team did not send the working notes. Therefore, their system could not be analyzed.
- **(5th) Jorge\_Owl**. This user did not send any working notes. Therefore, his system could not be analyzed.
- **(6th) INFOTEC-LaBD** (Cabrera-Pineda, S. Tellez, and Miranda, 2023). This team evaluated several UMAP features (Uniform Manifold Approximation and Projection) (McInnes et al., 2018), which incorporate a nonlinear dimensional reduction method of unigrams, bigrams, and 4-character-n-grams. Then, they evaluated SVMs classifiers with linear and non-linear kernels using five-fold stratified validation to evaluate different hyperparameters using Grid Search. They get similar position in the ranking for all the traits.
- **(7th) UMUTeam** (Pan, Caparrós-Laiz, and Almela, 2023). The UMUTeam employed a basic approach based on Transformers. Basically, they fine-tuned the MarIA model (Gutiérrez Fandiño et al., 2022) to create a classifier for each trait. They achieved a better position in the ranking for the psychographic traits than for the demographic traits.

- **(8th) NLP\_URJC** (Rodríguez-García, 2023). This team proposed a hybrid approach that evaluates different strategies for solving each task. These strategies involved MarIA (Gutiérrez Fandiño et al., 2022), BETO (Cañete et al., 2020), or traditional models such as Logistic Regression to identify the four traits.
- **(9th) Dataverse** (López-Ávila et al., 2023). Similar to other participants, the Dataverse team proposed an architecture based on BETO (Cañete et al., 2020) to accomplish the proposed task. They apply a preprocessing stage by removing emojis and transforming all tweets in lowercase. Note that this is a similar approach as the one conducted by the UMUteam, but replacing MarIA with BETO. This team improved on the proposed baseline.
- **(10th) UC3M** (Brandon et al., 2023). This team explored SVMs, Convolutional Neural Networks (CNNs) and two multilingual Transformers: multilingual BERT (Devlin et al., 2019) and XLM-T (Barbieri, Espinosa Anke, and Camacho-Collados, 2022). The classification report showed that they obtained limited recall in multi-classification problems for celebrities, when detecting the profession, and for the right and left wings, when detecting political ideology in multiclass. The limited recall caused that the UC3M team did not outperform the proposed baseline in the overall ranking. However, they achieve their best results in the gender prediction task, reaching 7th place in the ranking.
- **(11th) RandomSeed42**. These participants decided to withdraw their submission for the shared task. We keep the user in the leaderboard to maintain the ranking.

## 5 Results and discussion

The official leaderboard of the PoliticES 2023 shared task is shown in Table 2. It shows the results of the 11 participants plus a baseline based on Bag of Words and Logistic Regression. All traits are evaluated with the macro average F1-score and the final ranking is the average of the four traits. The

source code of this baseline is publicly available at <https://codalab.lisn.upaclay.fr/competitions/10173#participate>.

The system with the best overall performance is the one proposed by ELiRF-VRAIN with an average macro-f1 of 81.1319 using a voting scheme and Transformers. This result is followed by HiTZ-Ixa and ESCOM-IPN with an average macro-f1 of 79.3477 and 78.5280, respectively. It can be seen that the results are very competitive as most of the participants outperform the proposed baseline.

Regarding the results per trait, on the one hand, in terms of demographic characteristics, gender and profession have similar results, with the best score for gender being 82.9633% (ELiRF VRAIN) and for profession 86.0821% (HiTZ-Ixa). However, there are two labels for gender (male and female) but three for profession (politician, celebrity, and journalist), making gender the most difficult demographic to identify. For the psychographic trait, political ideology, the multi-class classification was the most complex, obtaining the best F1 score of 69.1334% (ESCOM-IPN), very similar to the second F1 score (69.1309%, ELiRF VRAIN). The binary trait of political ideology achieves higher scores, with 89.0715% in first place (ELiRF VRAIN) and 89.4368% in second place (ESCOM-IPN).

Regarding the approaches used, similar to the previous edition of PoliticES (García-Díaz et al., 2022), most teams proposed Transformers as the way to go, mainly fine-tuning the pre-trained models to identify the demographic and psychographic traits. The Transformers most widely used are BETO (Cañete et al., 2020) and MarIA (Gutiérrez Fandiño et al., 2022), but multilingual Transformers such as mBERT (Devlin et al., 2019), XLM-T (Barbieri, Espinosa Anke, and Camacho-Collados, 2022) were also evaluated. Only one team evaluated PolitiBETO (Villa-Cueva et al., 2022), released in the past edition of PoliticES, that was trained using the PoliticES 2022 dataset. Traditional models based on machine learning algorithms are also very popular and achieve good positions in the ranking. However, there are also participants promoting new ideas, such as the ideological discrepancy correction heuristic proposed by ELiRF-VRAIN, which aligns political ideology in case of discrepancies

Ranking			Traits			
#	Team	Ranking	Gender	Profession	BI	MI
01	ELiRF-VRain	81.1319	82.9633 (01)	82.7618 (03)	89.6715 (01)	69.1309 (02)
02	HiTZ-Ixa	79.3477	79.5627 (02)	86.0824 (01)	87.7588 (06)	63.9871 (06)
03	ESCOM-IPN	78.5280	76.9522 (03)	78.5898 (05)	89.4368 (02)	69.1334 (01)
04	INGEOTEC	77.7584	71.1549 (08)	83.7945 (02)	89.1394 (03)	66.9448 (03)
05	Jorge_Owl	77.1708	76.9522 (03)	76.7503 (06)	88.0666 (04)	66.9141 (04)
06	INFOTEC-LaBD	76.5324	74.3540 (04)	79.1357 (04)	87.9303 (05)	64.7094 (05)
07	UMUTEAM	69.2255	68.3632 (09)	61.6362 (09)	86.0017 (07)	60.9011 (07)
08	NLP_URJC	67.5665	72.8925 (05)	70.5932 (08)	72.2695 (10)	54.5109 (08)
09	Dataverse	66.6668	72.4442 (06)	73.1998 (07)	78.4927 (08)	42.5304 (09)
-	BASELINE	65.2679	66.3429 (-)	60.2390 (-)	79.7701 (-)	54.7196 (-)
10	UC3M	59.4392	71.4114 (07)	56.8049 (10)	73.3694 (09)	36.1709 (11)
11	RandomSeed42	53.8552	61.1954 (10)	54.3934 (11)	63.4900 (11)	36.3420 (10)

Table 2: PoliticES official rankings (rankings per trait are shown in parentheses). The traits are gender, profession, binary ideology (BI), and multiclass ideology (MI). The traits are evaluated using the macro-averaged F1-score and the final ranking is the average of all traits.

between binary and multiclass classification models.

## 6 Conclusions

This paper presents the second edition of the PoliticES task in IberLEF 2023. It is a cluster profiling task for political ideology in Spanish. Political ideology is a psychographic characteristic that can be used to understand individual and social behavior. Due to its relevance, we intend to promote the research of cluster profiling for political ideology in Spanish by organizing this collaborative task.

We are pleased with the impact of the PoliticES task, as 43 teams registered for it through CodaLab, which is still accessible at the following link: <https://codalab.lisn.upsaclay.fr/competitions/10173>. Finally, of all the registered participants, 11 submitted results and 8 presented a description of their proposal. As expected, approaches based on Transformers are the trend solutions, but some of them also used traditional machine learning systems or even a combination of both. Finally, it should be mentioned that multi-class political ideology is the most difficult feature for the participants to classify.

For future work, we plan to expand the dataset by including more clusters and increasing the diversity of professions. Another idea that has been proposed but not yet implemented is to add more subtasks to the author analysis. For example, we plan to add a stance detection subtask to determine

which authors are in favor of certain issues and which users are against them. We can use this information to define clusters of users and see if there is a relationship between issues and political ideology.

## Acknowledgements

This work is part of the research projects LaTe4PoliticES (PID2022-138099OB-I00) funded by MCIN/AEI/10.13039/501100011033 and the European Fund for Regional Development (FEDER)-a way to make Europe and LaTe4PSP (PID2019-107652RB-I00/AEI/10.13039/501100011033) funded by MCIN/AEI/10.13039/501100011033. This work is also part part of the research projects AIInFunds (PDC2021-121112-I00) and LT-SWM (TED2021-131167B-I00) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. It also has been partially supported by Project CONSENSO (PID2021-122263OB-C21), Project MODERATES (TED2021-130145B-I00) and Project SocialTox (PDC2022-133146-C21) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR, Project PRECOM (SUBV-00016) funded by the Ministry of Consumer Affairs of the Spanish Government, Project FedDAP (PID2020-116118GA-I00) supported by MICINN/AEI/10.13039/501100011033 and WeLee project (1380939, FEDER

Andalucía 2014-2020) funded by the Andalusian Regional Government. Salud María Jiménez-Zafra has been partially supported by a grant from Fondo Social Europeo and the Administration of the Junta de Andalucía (DOC\_01073).

## References

- Acosta-Pacheco, A.-M., D.-P. De-La-Cruz-Sierra, H. Gu, J.-M. Suárez-Bautista, J. Hernández-Espinoza, M.-G. Hernández-Lom, L.-R. Merino-Vázquez, and O. Juárez Gambino. 2023. Participation of ESCOM’s NLP group at PoliticES-IberLEF2023: Voting Ensemble and basic Machine Learning methods applied to political ideology. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023)*, CEUR-WS.org.
- Ahuir, V., L.-F. Hurtado, F. García-Granada, and Sanchis. 2023. ELiRF-VRAIN at PoliticES-IberLEF2023: Dealing with Long Texts in Transformer-based Systems for User Profiling. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023)*, CEUR-WS.org.
- Barbieri, F., L. Espinosa Anke, and J. Camacho-Collados. 2022. XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 258–266, Marseille, France, June. European Language Resources Association.
- Baumgaertner, B., J. E. Carlisle, and F. Justwan. 2018. The influence of political ideology and trust on willingness to vaccinate. *PloS one*, 13(1):e0191728.
- Brandon, S., M. M. Del-Toro Carballo, M. S. Arias Fernández, and I. Segura Bedmar. 2023. UC3M at PoliticEs 2023: Applying The Basics. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023)*, CEUR-WS.org.
- Cabrera-Pineda, H., E. S. Tellez, and S. Miranda. 2023. INFOTEC-LaBD at PoliticES-IberLEF2023: Explainable Non-Linear Low-Dimensional Projections. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023)*, CEUR-WS.org.
- Cañete, J., G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, and J. Pérez. 2020. Spanish pre-trained bert model and evaluation data. *Pml4dc at iclr*, 2020:1–10.
- Conneau, A., K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, É. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451.
- Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.
- Fatke, M. 2017. Personality traits and political ideology: A first global assessment. *Political Psychology*, 38(5):881–899.
- Fernandez de Landam, J. and R. Agerri. 2023. HiTZ-IXA at PoliticES-IberLEF2023: Document and Sentence Level Text Representations for Demographic Characteristics and Political Ideology Detection. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023)*, CEUR-WS.org.
- García-Díaz, J. A., Á. Almela, G. Alcaraz-Mármol, and R. Valencia-García. 2020. UMUCorpusClassifier: Compilation and evaluation of linguistic corpus for Natural Language Processing tasks. *Procesamiento del Lenguaje Natural*, 65(0):139–142.

- García-Díaz, J. A., R. Colomo-Palacios, and R. Valencia-García. 2022. Psychographic traits identification based on political ideology: An author analysis study on spanish politicians' tweets posted in 2020. *Future Generation Computer Systems*, 130:59–74.
- García-Díaz, J. A., S. M. Jiménez Zafra, M. T. Martín Valdivia, F. García-Sánchez, L. A. Ureña López, and R. Valencia García. 2022. Overview of politices 2022: Spanish author profiling for political ideology. *Procesamiento del Lenguaje Natural*.
- Gutiérrez Fandiño, A., J. Armengol Estapé, M. Pàmies, J. Llop Palao, J. Silveira Ocampo, C. Pio Carrino, C. Armentano Oller, C. Rodriguez Penagos, A. Gonzalez Agirre, and M. Villegas. 2022. MarIA: Spanish language models. *Procesamiento del Lenguaje Natural*, 68.
- He, P., J. Gao, and W. Chen. 2021. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. *arXiv preprint arXiv:2111.09543*.
- Jiménez-Zafra, S. M., F. Rangel, and M. Montes-y Gómez. 2023. Overview of IberLEF 2023: Natural Language Processing Challenges for Spanish and other Iberian Languages. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023)*, CEUR-WS.org.
- López-Ávila, P. E., A. B. García-Gutiérrez, P. A. Gallegos-Ávila, R. Aranda, and M. A. Álvarez Carmona. 2023. Dataverse at PoliticES-IberLEF2023. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023)*, CEUR-WS.org.
- McInnes, L., J. Healy, N. Saul, and L. Großberger. 2018. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861.
- Pan, R., C. Caparrós-Laiz, and A. Almela. 2023. UMUTeam at PoliticES-IberLEF2023: Evaluating Transformers for Detecting Political Ideology in Spanish Texts. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023)*, CEUR-WS.org.
- Rodríguez-García, M. A. 2023. URJC-Team at PoliticES-IberLEF2023: Political Ideology Detection Using Hybrid Architecture. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023)*, CEUR-WS.org.
- Sanh, V., L. Debut, J. Chaumond, and T. Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
- Verhulst, B., L. J. Eaves, and P. K. Hatemi. 2012. Correlation not causation: The relationship between personality traits and political ideologies. *American journal of political science*, 56(1):34–51.
- Villa-Cueva, E., I. González-Franco, F. Sanchez-Vega, and A. P. López-Monroy. 2022. Nlp-cimat at politices 2022: Politibeto, a domain-adapted transformer for multi-class political author profiling. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*. CEUR Workshop Proceedings, CEUR-WS, A Coruna, Spain.